

Linguistische Korpora im Spanischunterricht.

Eine didaktische Exploration.

Inaugural-Dissertation zur Erlangung der Doktorwürde
der Fakultät für Sprach-, Literatur- und Kulturwissenschaften der
Universität Regensburg

vorgelegt von

Nadine Patricia Meier

aus Saal an der Donau

2018

Regensburg

2018

Erstgutachterin: Apl. Prof. Dr. Annette Endruschat

Zweitgutachterin: Prof. Dr. Ingrid Neumann-Holzschuh

Dankesworte

Das Verfassen dieser Arbeit wäre ohne die Hilfe einiger Personen nicht möglich gewesen. Mein größter Dank gilt meiner Betreuerin, Frau apl. Prof. Dr. Annette Endruschat, die mir unermüdlich mit Rat und Tat zur Seite stand. Sie hatte stets ein offenes Ohr für mich und widmete meiner Arbeit viel Zeit. Ihre Anregungen, Hinweise und Ratschläge beeinflussten die vorliegende Dissertation in vielerlei Hinsicht.

Ebenso möchte ich Frau Prof. Dr. Ingrid Neumann-Holzschuh dafür danken, dass Sie sich dazu bereit erklärt hat, die Zweitkorrektur zu übernehmen. Mein Dank gilt außerdem den Teilnehmern des Linguistischen Oberseminars unter der Leitung von Frau Prof. Dr. Maria Selig und Frau Prof. Dr. Ingrid Neumann-Holzschuh. Die anregenden Diskussionen mit den Teilnehmern des Oberseminars führten zu einigen Überlegungen, welche die Arbeit in eine andere Richtung lenkten und um neue Facetten ergänzten.

Des Weiteren danke ich den Spanischschülern des Walter-Gropius-Gymnasiums Selb, die allen Unterrichtseinheiten, welche mithilfe von Korpora durchgeführt wurden, stets offen und motiviert gegenüberstanden. Ihre Rückmeldungen halfen mir in besonderem Maße bei der Erstellung und Verbesserung der Unterrichtsmaterialien.

Mein besonderer Dank gilt meinem Sohn Josef (,Seppi‘), der mir seit seiner Geburt zeigt, was im Leben wirklich wichtig ist, meinem Ehemann Josef, meinen Eltern Margit und Rudolf†, meiner Schwester Susanne, Josef (dem Lebensgefährten meiner Mutter), meinen Schwiegereltern Irmgard und Josef sowie allen meinen Freunden – insbesondere Nadine, Sabine und Stephanie – dafür, dass sie alle mich während meiner Promotionszeit moralisch unterstützt haben und immer für mich da waren.

Inhaltsverzeichnis

Einleitung: Aktuelle Anwendungsgebiete der Korpuslinguistik	1
1. Korpora und Korpuslinguistik	8
1.1. Was versteht man unter einem Korpus?	9
1.1.1. Korpora und Sprachnorm	15
1.1.2. Repräsentativität von Korpora – zwingende Eigenschaft oder unerreichbares Ziel?	26
1.2. Was ist Korpuslinguistik?	36
1.2.1. Entwicklung der Korpuslinguistik	40
1.2.2. Theoretische Linguistik und Korpuslinguistik	44
1.2.3. Korpuslinguistik: Methode oder Disziplin?	49
1.3. Ansätze der Korpuslinguistik	52
1.3.1. <i>Corpus-based approach</i> oder korpusgestützter Ansatz	52
1.3.2. <i>Corpus-driven approach</i> oder korpusbasierter Ansatz	54
1.3.2.1. Der korpusbasierte, quantitative Ansatz	55
1.3.2.2. Der korpusbasierte, quantitativ-qualitative Ansatz	56
1.4. Verfahren und Werkzeuge der Korpuslinguistik	57
1.4.1. <i>Type-token</i> -Relation und Frequenzlisten	58
1.4.2. Stichwortsuche – Suche nach Lexemen, Wortformen und Wortteilen	59
1.4.3. Konkordanzen und <i>Key Word in Context</i>	60
1.4.4. Suche mittels regulärer Ausdrücke	61
1.4.5. Annotation von Korpora: <i>tagging</i> und <i>parsing</i>	62
1.4.6. Abfragetools	66
1.4.7. Vergleichbarkeit von Daten: Normalisierung	67
1.4.8. Kollokationen und statistische Verfahren für deren Ermittlung	69
2. Kategorisierung der Korpora	72
2.1. Sprachbezug	73
2.1.1. Referenzkorpora	74
2.1.1.1. CREA – <i>Corpus de Referencia del Español Actual</i>	75
2.1.1.2. CORPES – <i>Corpus del Español del Siglo XXI</i>	79
2.1.1.3. CdE:Orig – Der aktuelle Teil des <i>Corpus del Español (Género/Histórico)</i>	82
2.1.2. Spezialkorpora	86
2.1.2.1. COLA – <i>Corpus Oral de Lenguaje Adolescente</i>	87
2.1.2.2. <i>Spanish in Texas Corpus</i>	89
2.1.3. Opportunistische Korpora	91
2.1.3.1. <i>Spa_news_2011_3M (FCS Aggregator)</i>	93
2.1.3.2. <i>Spa_newscrawl_2011</i> (Wortschatz Universität Leipzig)	93

2.2. Funktionalität.....	95
2.2.1. Historische Korpora.....	95
2.2.1.1. CORDE – <i>Corpus Diacrónico del Español</i>	96
2.2.1.2. CDH – <i>Corpus del Nuevo Diccionario Histórico del Español</i>	100
2.2.1.3. CdE:Orig (Gesamt) – <i>Corpus del Español (Género/Histórico)</i>	101
2.2.1.4. <i>Biblia Medieval</i>	103
2.2.2. Lernerkorpora	105
2.2.2.1. CORELE – <i>Corpus Oral de Español como Lengua Extranjera (ELE)</i>	108
2.2.2.2. SPLLOC – <i>Spanish Learner Language Oral Corpora</i>	110
2.2.2.3. <i>Fono.ele</i>	111
2.2.2.4. LANGSNAP – <i>Languages and Social Networks Abroad</i>	113
2.2.2.5. CAES – <i>Corpus de aprendices de español como lengua extranjera</i> ..	114
2.2.2.6. CEDEL2 – <i>Corpus Escrito del Español L2</i>	116
2.3. Sprachenauswahl	118
2.3.1. <i>Corpus Trilingüe Paralelo Grial</i>	120
2.3.2. CLUVI – <i>Corpus Lingüístico da Universidade de Vigo</i>	121
2.4. Medium.....	123
2.4.1. Korpora geschriebener Sprache.....	124
2.4.1.1. CEA – <i>Corpus del Español Actual</i>	125
2.4.1.2. <i>Corpus periodístico Hemero</i>	126
2.4.2. Korpora gesprochener Sprache.....	128
2.4.2.1. <i>Macrocorpus de la norma lingüística culta de las principales ciudades del mundo hispánico</i>	130
2.4.2.2. C-ORAL-ROM.....	131
2.4.2.3. CORLEC – <i>Corpus Oral de Referencia de la Lengua Española Contemporánea</i>	134
2.4.2.4. C-Or-DiAL – <i>Corpus Oral Didáctico Anotado Lingüísticamente</i>	135
2.4.3. Multimodale Korpora	137
2.4.3.1. <i>Sacodeyl Search Tool</i>	138
2.4.3.2. <i>Backbone Spanish</i>	139
2.4.3.3. SPT – <i>Spanish Corpus Proficiency Level Training</i>	141
2.5. Annotation	142
2.5.1. Morphosyntaktische Annotation.....	142
2.5.1.1. CT-IULA – <i>Corpus Tècnic del IULA</i>	143
2.5.1.2. <i>El Grial</i>	145
2.5.2. Syntaktische Annotation.....	147
2.5.2.1. BDS – <i>Base de Datos Sintácticos del Español Actual</i>	148
2.5.2.2. <i>Europarl-es</i>	149
2.5.2.3. AnCora.....	150
2.5.3. Semantische Annotation	152
2.5.3.1. <i>SenSem (Español)</i>	152
2.5.3.2. <i>Internet with semantic roles (VISL)</i>	153

2.6. Größe	154
2.6.1. CdE:Nuevo – <i>Corpus del Español (Web/Dialectos)</i>	157
2.6.2. <i>Google Books: Spanish</i>	163
2.6.3. <i>NOW Corpus: Spanish</i>	164
2.7. Persistenz	167
2.8. Verfügbarkeit.....	169
2.8.1. Im Internet zugängliche Korpora.....	170
2.8.1.1. <i>esTenTen</i>	170
2.8.1.2. ADMYTE – <i>Archivo Digital de Manuscritos y Textos Españoles</i>	173
2.8.2. Als CD-ROM verfügbare Korpora	174
2.8.2.1. PRESEEA – <i>Proyecto para el Estudio Sociolingüístico del Español de España y de América</i>	175
2.8.2.2. LEXESP – <i>Léxico informatizado del español</i>	176
2.9. Das Internet – Ein Korpus?	178
3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht	192
3.1. Lerntheoretischer Hintergrund	197
3.1.1. Ansätze und Methoden des Fremdsprachenunterrichts	197
3.1.2. Der Gemeinsame Europäische Referenzrahmen für Sprachen.....	209
3.1.3. Der Lehrplan für Spanisch an bayerischen Gymnasien.....	212
3.1.3.1. Sprache	216
3.1.3.2. Umgang mit Texten und Medien.....	219
3.1.3.3. Interkulturelles Lernen und Landeskunde	220
3.1.3.4. Lernstrategien und Methoden selbstständigen Arbeitens	222
3.2. Motive für die Verwendung von Korpora im Unterricht	223
3.2.1. Authentisch, genuin, real oder realistisch – der Begriff der sogenannten Authentizität	223
3.2.2. Der Begriff der kommunikativen Kompetenz	232
3.2.3. Der Begriff der Lernerautonomie	238
3.3. Einsatzmöglichkeiten von Korpora im Fremdsprachenunterricht.....	245
3.3.1. Direkte Nutzung von Korpora	246
3.3.1.1. <i>Exploiting to teach</i>	248
3.3.1.2. <i>Teaching to exploit</i>	257
3.3.2. Indirekte Nutzung von Korpora.....	267
3.3.2.1. Beitrag für die Erstellung von Wörterbüchern	267
3.3.2.2. Entwicklung von Lehrplänen und Unterrichtsmaterialien.....	281
3.3.2.2.1. Aktuelle Lehrpläne und Unterrichtsmaterialien	284
3.3.2.2.2. Vergleich einer Studie zum Ausdruck der Zukünftigkeit mit aktuellen Schulgrammatiken	297
3.3.2.2.3. Studie zur Auswahl des Vokabulars für Lehrwerke	301
3.3.2.3. Erstellung von Sprachtests.....	306
3.3.3. Unterrichtsorientierte Korpora	308

4. Nutzung von Korpora in der fremdsprachlichen Unterrichtspraxis	323
4.1. Lexik	323
4.1.1. Kollokationen	323
4.1.1.1. Kollokationen in Wörterbüchern	339
4.1.1.2. Kollokationen in Korpora	347
4.1.1.3. Erstellung von Unterrichtsmaterial zum Thema Kollokationen	351
4.1.1.3.1. Erweiterung der Ausdrucksfähigkeit anhand von Kollokationen	352
4.1.1.3.2. Erstellung einer Mindmap mithilfe von Korpora	358
4.1.1.3.3. Erarbeitung eines Kollokationsfelds mithilfe von Korpora	363
4.1.1.3.4. Erarbeitung von Nomen-Adjektiv-Kollokationen	369
4.1.1.3.5. Integration von neuem Vokabular in einen größeren sprachlichen Kontext	371
4.1.2. Phraseologismen	374
4.1.3. Synonyme	378
4.1.3.1. Ermittlung von Synonymen anhand des CdE	381
4.1.3.2. Erstellung eines Wortfelds anhand des CdE	383
4.1.3.3. Nutzung des CdE als Thesaurus	385
4.1.3.4. Verbesserung der Ausdrucksfähigkeit anhand von Synonymen	386
4.1.3.5. Klärung von Bedeutungs- und Verwendungsunterschieden	389
4.1.4. Homonyme und Polyseme	390
4.1.5. Wortbildung	397
4.1.5.1. Präfigierung und Suffigierung	398
4.1.5.2. Wortfamilien	399
4.1.6. Äquivalente des deutschen „werden“	401
4.1.7. <i>Ser, estar</i> und <i>hay</i>	405
4.2. Grammatik	407
4.2.1. Präpositionen	408
4.2.2. Konjugation regelmäßiger und unregelmäßiger Verben	410
4.2.2.1. <i>Presente de indicativo</i>	411
4.2.2.2. <i>Pretérito indefinido</i>	412
4.2.2.3. <i>Participio de pasado</i>	414
4.2.2.4. Allgemeine Nutzung von Korpora zur Ermittlung unregelmäßiger Verbformen	415
4.2.3. Relativsätze	416
4.2.4. Bedingungssätze	419
4.2.5. <i>Subjuntivo</i>	421
4.2.6. Fehlerkorrektur	427
4.3. Aussprache	429
4.4. Textverständnis	433
4.4.1. Hörverstehen	433
4.4.2. Leseverstehen	441

4.5. Landeskunde	447
4.5.1. Alltag und Hobbies spanischer Jugendlicher.....	447
4.5.2. Umgangssprache und Jugendsprache	449
4.5.3. Argentinisches Spanisch.....	452
4.5.4. Mehrsprachigkeit in Spanien	455
4.6. Eignung der bisher verfügbaren spanischen Korpora für den Unterricht.....	458
4.6.1. Referenzkorpora: CREA, CORPES und CdE:Orig	458
4.6.2. Spezialkorpora: COLA und <i>Spanish in Texas Corpus</i>	466
4.6.3. Opportunistische Korpora: <i>esTenTen</i> , <i>Spa_news_2011_3M</i> , <i>Spa_newscrawl_2011</i> , <i>Hemero</i> , CEA, CdE:Nuevo, <i>Google Books: Spanish</i> und <i>NOW Corpus: Spanish</i>	468
4.6.4. Historische Korpora: CORDE, CDH, CdE:Orig (Gesamt), <i>Biblia Medieval</i> und ADMYTE	472
4.6.5. Lernerkorpora: CORELE, SPLLOC, LANGSNAP, <i>Fono.ele</i> , CAES, CEDEL2 und SPT.....	474
4.6.6. Multilinguale Korpora: CLUVI und <i>Corpus Trilingüe Paralelo Grial</i>	476
4.6.7. Korpora gesprochener Sprache: <i>Macrocorpus</i> , C-ORAL-ROM, PRESEEA, CORLEC und C-Or-DiAL.....	477
4.6.8. Multimodale Korpora: SPT, <i>Sacodeyl</i> und <i>Backbone Spanish</i>	479
4.6.9. Annotierte Korpora.....	481
4.6.9.1. Morphosyntaktisch annotierte Korpora: CT-IULA und <i>El Grial</i>	481
4.6.9.2. Syntaktisch annotierte Korpora: BDS, <i>Europarl-es</i> und AnCora	483
4.6.9.3. Semantisch annotierte Korpora: <i>SenSem</i> und <i>Internet with Semantic</i> <i>Roles</i>	484
4.6.10. Korpora auf CD-ROM: PRESEEA, LEXESP, C-ORAL-ROM und <i>Macrocorpus</i>	485
4.7. Exkurs: Eigene Erstellung eines Korpus	487
4.7.1. Einführung in das Konkordanzprogramm <i>AntConc</i>	488
4.7.2. Funktionen des Konkordanzprogramms <i>AntConc</i>	489
4.7.3. Anwendungsbeispiele für die Nutzung eines selbst erstellten Korpus.....	505
4.7.3.1. Erstellung eigener Stoplisten im Programm <i>AntConc</i>	505
4.7.3.2. Erstellung von Grammatikübungen anhand bereits bekannter Texte..	507
4.7.3.3. Erstellung eines Korpus mit Schülern	508
5. Schlussfolgerungen: Möglichkeiten und Grenzen der Verwendung von Korpora für den Fremdsprachenunterricht.....	512
6. Bibliographie	531
6.1. Nachschlagewerke	531
6.2. Sekundärliteratur	533
6.3. Internetquellen	553
6.4. Schriftlicher und telefonischer Kontakt.....	575

7.	Abbildungsverzeichnis	576
8.	Anhang.....	582

Einleitung: Aktuelle Anwendungsgebiete der Korpuslinguistik

Der Begriff Korpuslinguistik ist in der aktuellen sprachwissenschaftlichen Forschung allgegenwärtig. Dabei stellt sich die Frage, in welchen Bereichen und wofür Korpora derzeit konkret genutzt werden. Korpora können auf den verschiedensten Gebieten angewandt werden, so beispielsweise bei der Sprachforschung, in verschiedenen Unterbereichen der Linguistik wie Lexik, Grammatik, Semantik, Pragmatik und Diskursanalyse, in der Soziolinguistik, der Stilistik, der Sprach- und Linguistiklehre, der historischen Linguistik, der Dialektologie, der Psycholinguistik, der Kulturwissenschaft sowie für die Sozialpsychologie.¹ Zudem hat die Korpuslinguistik neben anderen linguistischen Strömungen den Weg für die relativ junge Dokumentationslinguistik gebahnt, deren Ziel es ist, natürliche, insbesondere bedrohte Sprachen möglichst umfassend und repräsentativ zu dokumentieren.²

Lemnitzer/Zinsmeister geben an, dass die Korpuslinguistik in der Praxis vor allem für die Bereiche Orthographie, Wortbildung, Syntax, Computerlinguistik, Lexikologie und Lexikographie, Textlinguistik sowie Fremdsprachenerwerb und -vermittlung genutzt wird. In der Orthographie können mittels der Korpora Fragen zum tatsächlichen Sprachgebrauch, zu den quantitativen Verhältnissen verschiedener Varianten, zu orthographischen Unsicherheiten sowie zu Besonderheiten bestimmter Textsorten geklärt werden. Im Bereich der Wortbildung dienen sie als Prüfstein für theoretische Annahmen zu Regeln, Regularitäten und Beschränkungen in Wortbildungsverfahren sowie zur Überprüfung der Produktivität bestimmter Wortbildungsmuster. Anhand von Korpora wird auf dem Gebiet der Syntax unter anderem nach Beispielen im Rahmen einer bestimmten Theorie gesucht oder es werden Frequenzdaten zu bestimmten Phänomenen erhoben. In der Computerlinguistik dienen Korpora als Datenquelle für das empirische Arbeiten. Zudem werden sie zum Training statistischer Programme genutzt. Eine besondere Rolle spielt die Existenz von Korpora im Bereich der Lexikologie und

¹ Vgl. McEnery, Tony/ Wilson, Andrew (1996): *Corpus Linguistics*. Edinburgh: Edinburgh University Press.

² Vgl. Endruschat, Annette/ Ferreira, Vera (2011): "Dokumentationslinguistik – Ein Forschungsüberblick". In: Endruschat, A./ Ferreira, V. (Hrsg.) (2011): *Sprachdokumentation und Korpuslinguistik – Forschungsstand und Anwendung: Akten des 8. Deutschen Lusitanistentags*. München: Meidenbauer, S. 1, 5.

Lexikographie. Sie können wichtige Hinweise zur Auswahl der Lexeme geben, dienen als Informationsquelle auf allen linguistischen Ebenen sowie als Quelle für Verwendungsbeispiele, sie geben Auskunft über die Nutzungshäufigkeit und über typische sprachliche Kontexte.³

Laut Parodi beschäftigt sich der Großteil der derzeit veröffentlichten Artikel und Bücher in Bezug auf sprachliche Korpora jedoch mit der englischen Sprache.⁴ Wie die im Folgenden aufgeführten Studien zeigen, werden Korpora allerdings auch in der spanischsprachigen Forschung bereits auf einigen der erwähnten Gebiete angewandt.

Im Bereich der Nutzung von Korpora für die historische Linguistik wurde im Jahr 2009 eine Sammlung von 21 Studien zur diachronen Linguistik veröffentlicht, welche auf dem *Coloquio Internacional sobre Corpus Diacrónicos en Lenguas Iberorromances* im Jahr 2007 vorgestellt wurden.⁵ Außerdem ist im Jahr 2016 das Buch *Lingüística de corpus y lingüística histórica iberorrománica* erschienen, in dem aktuelle Tendenzen der iberoromanischen historischen Linguistik besprochen sowie anhand historischer Korpora des Spanischen, Portugiesischen, Katalanischen und Galicischen durchgeführte Untersuchungen beschrieben werden. Es finden sich hier beispielsweise Beiträge über die Entstehung des Buchstaben ‚J‘ als konsonantische Schreibung oder die Variation des kolonialen Spanisch.⁶

Im Hinblick auf die Übersetzungswissenschaften ist eine Analyse der englischen Übersetzung phraseologischer Einheiten in *Cien Años de Soledad* zu nennen, welche im englischen Sammelband *Corpus Data across Languages and Disciplines* als einziger Aufsatz zur spanischen Sprache veröffentlicht wurde.⁷

³ Vgl. Lemnitzer, Lothar/ Zinsmeister, Heike (2010): *Korpuslinguistik. Eine Einführung*. Tübingen: Narr Francke Attempto Verlag, S. 2, 125-141.

⁴ Vgl. Parodi, Giovanni (2007a): „Introduction. Catching up with corpus linguistics: Register-diversified studies from different corpora in different Spanish-speaking countries“. In: Parodi, G. (Hrsg.) (2007): *Working With Spanish Corpora*. London: Continuum, S. 2.

⁵ Vgl. Enrique-Arias, Andrés (Hrsg.) (2009): *Diacronía de las lenguas iberorrománicas: nuevas aportaciones desde la lingüística de corpus*. Madrid: Iberorrománica.

⁶ Vgl. Kabatek, Johannes (Hrsg.) (2016): *Lingüística de corpus y lingüística histórica iberorrománica*. Berlin/ Boston: Walter De Gruyter.

⁷ Vgl. Quinteiro, Natalia/ Quiroz, Gabriel (2012): „Analysis of the Translation of the Phraseological Units of *Cien Años de Soledad* into English: A Corpus-based Study“. In: Pezik, P. (Hrsg.) (2012): *Corpus Data across Languages and Disciplines*. Frankfurt am Main [u.a.]: Lang, S. 245-254.

1. Korpora und Korpuslinguistik

Eine Studie zur Erstellung eines Webkorpus mit automatisch annotierten semantischen Rollen von Valverde-Ibáñez/Bick⁸ ist für die Semantik aufzuführen und im Bereich der Phonologie wurde eine Abhandlung zum englisch-spanischen *Code-Switching* in der Facebook-Kommunikation von Autoren aus Arkansas durchgeführt.⁹

Der Aufsatz „Evidentielle Markierung von fremden Äußerungen. Eine Analyse des spanischen Imperfekts in journalistischen Texten“ von Verónica Böhm aus dem Jahr 2013 ist für den grammatikalischen Bereich zu nennen.¹⁰ Außerdem wurde von Anja Hennemann eine Studie zum Gebrauch des peninsularen *pretérito perfecto* zur Beschreibung abgeschlossener Handlungen durchgeführt.¹¹ In Bezug auf die Syntax wurde beispielsweise der Aufsatz „Towards a subject cartography in Catalan: the example of oblique subjects“ von Vega im Jahr 2013 veröffentlicht.¹²

Auf dem Gebiet der Textlinguistik kann das Buch *Lingüística de corpus: de la teoría a la empiria* von Parodi genannt werden, welches im Jahr 2010 erschienen ist. Darin beschreibt der Autor neben allgemeinen Informationen, Definitionen und Vorgehensweisen zur Korpuslinguistik auch die Benutzeroberfläche des spanischen Korpus *El Grial*, auf welches in Kapitel 2.5.1.2. näher eingegangen werden wird, sowie qualitative und quantitative Untersuchungen zu verschiedenen Registern und Textgenres.¹³

Der Aufsatz „Variable development of intersubjectivity in Spanish“ von Rivas/Brown aus dem Jahr 2010 kann im Hinblick auf die Diskursanalyse aufgeführt

⁸ Vgl. Valverde-Ibáñez, Pilar M./ Bick, Eckhard (2010): „A web corpus of Spanish automatically annotated with semantic roles“. In: Sánchez, A./ Almela, M. (Hrsg.) (2010): *A Mosaic of Corpus Linguistics: Selected Approaches*. Frankfurt am Main: Lang, S. 249-268.

⁹ Vgl. Ammermann, Anne (2013): „English-Spanish code-switching in the Facebook communication of Arkansas authors“. In: Hennemann, A./ Schlaak, C. (Hrsg.) (2013): *Korpuslinguistische Untersuchungen: Analysen einzelsprachlicher Phänomene*. Berlin: Frank & Timme, S. 187-198.

¹⁰ Vgl. Böhm, Verónica (2013): „Evidentielle Markierung von fremden Äußerungen. Eine Analyse des spanischen Imperfekts in journalistischen Texten“. In: Hennemann, A./ Schlaak, C. (Hrsg.) (2013): *Korpuslinguistische Untersuchungen: Analysen einzelsprachlicher Phänomene*. Berlin: Frank & Timme, S. 119-130.

¹¹ Vgl. Hennemann, Anja (2013): „*Ayer he hecho* vs. *ayer yo hice*. On the increasing use of the (Peninsular) Spanish Present Perfect to describe completed actions“. In: Hennemann, A./ Schlaak, C. (Hrsg.) (2013): *Korpuslinguistische Untersuchungen: Analysen einzelsprachlicher Phänomene*. Berlin: Frank & Timme, S. 101-118.

¹² Vgl. Vega, Jorge (2013): „Towards a subject cartography in Catalan: the example of oblique subjects“. In: Hennemann, A./ Schlaak, C. (Hrsg.) (2013): *Korpuslinguistische Untersuchungen: Analysen einzelsprachlicher Phänomene*. Berlin: Frank & Timme, S. 89-100.

¹³ Vgl. Parodi, Giovanni (2010): *Lingüística de corpus: de la teoría a la empiria*. Madrid: Iberoamericana.

werden.¹⁴ Außerdem fällt die Studie von Zulaica Hernández in diesen Bereich. Er untersuchte die Spanne zwischen anaphorisch gebrauchten Demonstrativa und deren vorausgehenden Referenten gemessen an der Anzahl von Sätzen.¹⁵ Ebenfalls in diesem Bereich ist die Untersuchung lexikalischer Bündel in gesprochener und geschriebener Sprache von Tracy-Ventura/Cortes/Biber angesiedelt. Sie analysierten die Funktionen sowie die Verteilung lexikalischer Bündel in einem großen spanischen Korpus mit Texten aus formellen Konversationen und akademischer Prosa. Zudem wurden die lexikalischen Bündel des Spanischen mit denjenigen des Englischen und anderer Sprachen verglichen.¹⁶

Unter den genannten Anwendungsgebieten sind Scherer zufolge die Erforschung sprachlicher Strukturen und Varietäten, die Erstellung von Wörterbüchern und Grammatiken, der Fremdsprachenunterricht, die Übersetzung sowie die Computerlinguistik die bedeutendsten.¹⁷ McCullough und Tagnin sehen jeweils drei Hauptanwendungsbereiche, zu denen bei beiden die Erstellung von Wörterbüchern sowie die Übersetzung gehören. Während McCullough allerdings die Forschung auf akademischem Niveau, beispielsweise in den Fachrichtungen Morphologie, Lexik, Syntax oder Pragmatik, als dritten Zweig nennt, gibt Tagnin den Fremdsprachenunterricht an.¹⁸

Für die sprachwissenschaftliche Forschung stellen Korpora einen unmittelbaren Zugriff auf reale Sprachdaten dar. Durch Korpora ist es Linguisten¹⁹ möglich, bereits

¹⁴ Vgl. Rivas, Javier/ Brown, Esther (2010): „Variable development of intersubjectivity in Spanish“. In: Sánchez, A./ Almela, M. (Hrsg.) (2010): *A Mosaic of Corpus Linguistics: Selected Approaches*. Frankfurt am Main: Lang, S. 61-78.

¹⁵ Vgl. Zulaica Hernández, Iker (2010): „Testing accessibility and givenness in a corpus of Spanish“. In: Sánchez, A./ Almela, M. (Hrsg.) (2010): *A Mosaic of Corpus Linguistics: Selected Approaches*. Frankfurt am Main: Lang, S. 99-118.

¹⁶ Vgl. Tracy-Ventura, Nicole/ Cortes, Viviana/ Biber, Douglas (2007): „Lexical bundles in speech and writing“. In: Parodi, G. (Hrsg.) (2007): *Working With Spanish Corpora*. London: Continuum, S. 217-231.

¹⁷ Vgl. Scherer, Carmen (2006): *Korpuslinguistik*. Heidelberg: Universitätsverlag Winter, S. 10-15.

¹⁸ Vgl. McCullough, James L. (2001): „Los usos de los corpórea de textos en la enseñanza de lenguas“. In: Trenchs Parera, M. (Hrsg.) (2001): *Nuevas Tecnologías para el Autoaprendizaje y la Didáctica de Lenguas*. Lleida: Milenio, S. 128. Und: Tagnin, Stella E.O. (2003): „Comet – A multilingual corpus for teaching and translation“. In: Lewandowska-Tomaszczyk, B. (Hrsg.) (2003): *PALC 2001: Practical Applications in Language Corpora*. Frankfurt am Main: Lang, S. 536.

¹⁹ In der vorliegenden Arbeit werden aus Gründen der Lesbarkeit männliche Personenbezeichnungen für beide Geschlechter verwendet. Dabei sind in jedem Fall männliche und weibliche Linguisten, Lehrer, Schüler, Lerner etc. gemeint, soweit nicht anders angegeben.

aufgestellte Theorien empirisch zu prüfen und anhand der Daten neue Hypothesen aufzustellen.²⁰ In der Sprachwissenschaft werden elektronische Korpora schon seit den 1960er Jahren genutzt, doch in der Sprachlehre werden sie erst seit kurzer Zeit eingesetzt.²¹ Laut Siepmann sind Korpora aus der Linguistik und der (Lerner-) Lexikographie heute nicht mehr wegzudenken. Auch in der universitären Lehre, beispielsweise in Übersetzungskursen sowie in sprachwissenschaftlichen Seminaren, werden sie zunehmend verwendet, doch im fremdsprachlichen Unterricht finden sie seines Wissens nach bisher kaum Anwendung.²²

Es stellt sich also die Frage, ob und inwieweit Korpora bereits im Fremdsprachenunterricht genutzt werden. Um einen Einblick in die aktuelle Situation der Verwendung von Korpora im Spanischunterricht zu bekommen, wurde daher eine Umfrage unter Spanischlehrern durchgeführt. Es handelt sich dabei um keine groß angelegte, repräsentative Studie, sondern lediglich um eine kleine Umfrage, die dennoch einen guten Überblick über den Ist-Zustand gibt und auch bezüglich der Angaben zur Computernutzung als realistisch eingestuft werden kann. Hätten lediglich junge Lehrkräfte im Alter von 24 bis 35 Jahren an der Umfrage teilgenommen oder würden viele teilnehmende Lehrkräfte den Computer generell nur wenig nutzen, so wären die Angaben zur Verwendung digitaler Korpora wenig aussagekräftig.

Die Umfrage zur derzeitigen Nutzung linguistischer Korpora im Spanischunterricht, an der 32 Spanischlehrkräfte teilnahmen, wurde im Sommer 2016 durchgeführt. Der verwendete Fragebogen befindet sich in Anhang 1. Fast die Hälfte der Lehrkräfte ist aus Bayern, es nahmen allerdings auch einige Lehrer aus Schleswig-Holstein, Nordrhein-Westfalen, Niedersachsen, Bremen und Berlin teil. Ein Großteil der Teilnehmer unterrichtet an einem Gymnasium. Je ein gutes Drittel der Lehrkräfte zählt zur Altersgruppe der 24- bis 35-Jährigen oder der 36- bis 45-Jährigen. Die restlichen Teilnehmer sind zwischen 46 und 65 Jahren alt. Dreiviertel der Befragten sind weiblich,

²⁰ Vgl. Scherer (2006): op. cit., S. 15.

²¹ Vgl. Braun, Sabine/ Chambers, Angela (2009): „Elektronische Korpora als Ressource für den Fremdsprachenunterricht“. In: Jung, U. O. H. (Hrsg.) (2009): *Praktische Handreichung für Fremdsprachenlehrer*. Frankfurt am Main [u.a.]: Lang, S. 330.

²² Vgl. Siepmann, Dirk (2009): „Was der Fremdsprachenunterricht von der Korpuslinguistik erwarten darf“. In: Jung, U. O. H. (Hrsg.) (2009): *Praktische Handreichung für Fremdsprachenlehrer*. Lang: Frankfurt am Main [u.a.], S. 328.

was durch die Tatsache erklärt werden kann, dass es generell mehr weibliche als männliche Lehrkräfte im Bereich der Fremdsprachen gibt. Von über zwei Drittel der Teilnehmer wird der Computer für rund 70% der Unterrichtsvorbereitung verwendet, wobei 64,5% angeben, ihn kaum im Unterricht mit den Schülern zu nutzen.

Keiner der Teilnehmer hat bisher ein Korpus zur Vor- oder Nachbereitung des Unterrichts verwendet, wobei 90% als Gründe mangelnde Kenntnis der existierenden Korpora anführen, 44% mangelnde Kenntnis der Funktionsweise derselben, 19% mangelnde Zeit für die Einarbeitung in die Korpusthematik und 9% mangelnde Zeit für die Arbeit mit Korpora. Mehrfachnennungen waren möglich. Insgesamt geben die befragten Lehrkräfte die Wahrscheinlichkeit dafür, dass sie ein Korpus nutzen würden, sofern sie eine Art Leitfaden hätten, auf einer Skala von 1 bis 6 mit 2,8 an. Knapp die Hälfte der Teilnehmer kann es sich gut vorstellen, ein Korpus zu nutzen, zwölf Personen vielleicht und lediglich bei fünf Lehrkräften ist eine Verwendung von Korpora unwahrscheinlich.

Auch im Unterricht wurden Korpora bisher von keinem der Befragten genutzt. In diesem Bereich führen 94% als Grund die mangelnde Kenntnis der existierenden Korpora an, 56% mangelnde Kenntnis der Funktionsweise ihrerseits und 25% seitens der Schüler, 19% mangelnde Zeit für die Einarbeitung in die Korpusthematik ihrerseits und 6% seitens der Schüler, 12,5% mangelnde Zeit für die Arbeit mit Korpora ihrerseits sowie 12,5% mangelnde Unterrichtszeit. Auch hier konnten mehrere Gründe angegeben werden. Für die Nutzung von Korpora im Unterricht geben die Lehrer die Wahrscheinlichkeit dafür, dass sie ein Korpus im Unterricht verwenden würden, sofern sie eine Art Leitfaden hätten, auf einer Skala von 1 bis 6 mit 3,3 an. In diesem Bereich können sich elf Personen vorstellen, ein Korpus im Unterricht mit Schülern zu nutzen, 13 Lehrkräfte eventuell und 6 Teilnehmer sehen diese Möglichkeit als sehr unwahrscheinlich an.

Keiner der Teilnehmer hat bislang ein eigenes Korpus erstellt, was aufgrund der bisherigen Umfrageergebnisse nicht überrascht. Hier geben 97% mangelndes Wissen über die Konstruktion eines Korpus an, 41% mangelnde Zeit für die Erstellung desselben und 19% mangelnde Unterrichtszeit. Die Wahrscheinlichkeit der eigenen Entwicklung eines Korpus betitelten sie mit 3,6 auf einer Skala von 1 bis 6. Immerhin ist die Erstellung eines Korpus bei 8 Personen wahrscheinlich, 13 Lehrkräfte würden eventuell ein Korpus erstellen und zehn von ihnen können es sich gar nicht vorstellen. Die vollständigen

1. Korpora und Korpuslinguistik

Umfrageergebnisse sind in Anhang 2 aufgelistet.

Das Resultat der Umfrage zur derzeitigen Nutzung von Korpora im Unterricht legt somit nahe, dass die befragten Spanischlehrer die bereits bestehenden Korpora durchaus zur Vor- und Nachbereitung ihres Unterrichts oder sogar in ihrem Unterricht mit den Schülern zusammen nutzen würden, sie jedoch wenig über Korpora, deren Funktionsweise und die Korpuslinguistik an sich wissen. Zudem spielt die Zeit eine wichtige Rolle, da die Einarbeitung, aber auch die Korpusarbeit an sich, Zeit kosten, und diese aufgrund der vollgepackten Lehrpläne bei den ohnehin vielbeschäftigten Lehrkräften rar gesät ist.

Aufgrund der Tatsache, dass wohl generell der Großteil der Lehrkräfte noch nie mit einem Korpus gearbeitet hat und einige von ihnen schlechthin nicht wissen, worum es sich bei der Korpuslinguistik handelt, soll die folgende Arbeit ermitteln, wie linguistische Korpora im und für den Spanischunterricht genutzt werden können. Es stellt sich die Frage, ob Korpora sinnvoll in den Unterricht integriert werden können oder ob sie zu komplex sind und daher der wissenschaftlichen Forschung vorbehalten bleiben sollten.

Zunächst sollen im ersten Kapitel relevante Begriffe wie ‚Korpus‘ und ‚Korpuslinguistik‘ geklärt werden, wobei hinsichtlich der Korpora auf das Verhältnis zur Sprachnorm sowie auf die Repräsentativität des gewählten Sprachausschnitts näher eingegangen wird. In Bezug auf den Begriff ‚Korpuslinguistik‘ sind die Entwicklung derselben, das Verhältnis zur theoretischen Linguistik sowie der Status derselben als Methode oder Disziplin relevant. Schließlich werden im ersten Kapitel noch Ansätze, Verfahren und Werkzeuge der Korpuslinguistik beschrieben.

Das zweite Kapitel beschäftigt sich mit der Kategorisierung der Korpora, wobei im Zuge dieser Einteilung die derzeit bekannten Korpora der spanischen Sprache den jeweils zutreffenden Kategorien zugeordnet werden. Korpora werden generell nach Sprachbezug, Funktionalität, Sprachenauswahl, Medium, Annotation, Größe, Verfügbarkeit und Persistenz (hinsichtlich der Stabilität der Daten) eingeteilt²³, wobei ein Korpus stets mehreren Kategorien gleichzeitig angehört. Zum Schluss dieses Kapitels wird schließlich der Frage nachgegangen, ob das Internet ein Korpus darstellt.

²³ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 102-107.

Die theoretischen Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht werden im dritten Kapitel näher durchleuchtet. Dabei wird zunächst auf Ansätze und Methoden des Fremdsprachenunterrichts, den *Gemeinsamen Europäischen Referenzrahmen für Sprachen* (GER) sowie exemplarisch auf den Lehrplan für Spanisch an bayerischen Gymnasien eingegangen. Daraufhin werden die Begriffe Authentizität, kommunikative Kompetenz und Lernerautonomie als Motive für die Nutzung von Korpora im Unterricht näher erläutert, bevor schließlich die unterschiedlichen Einsatzmöglichkeiten von Korpora im Fremdsprachenunterricht veranschaulicht werden.

Das vierte Kapitel beschäftigt sich mit verschiedenen Möglichkeiten der Nutzung von Korpora im Unterricht, wobei das Hauptaugenmerk auf die Lexik, insbesondere auf die Arbeit mit Kollokationen und Synonymen, gelegt wird. Doch auch in anderen Bereichen des schulischen Fremdsprachenunterrichts, wie der Grammatik, der Aussprache, dem Lese- oder Hörverstehen und sogar in der Vermittlung landeskundlicher Inhalte können Korpora angewandt werden. Des Weiteren wird in einem Unterkapitel zusammengefasst, welche Korpora aus dem zweiten Kapitel tatsächlich im und für den Spanischunterricht genutzt werden können, und für welche Bereiche des fremdsprachlichen Unterrichts sie sich eignen. Zum Schluss dieses sehr praxisorientierten Kapitels wird schließlich beschrieben, wie anhand des Programms *AntConc* aus den Lektionen eines Lehrbuchs sowie weiterer im Unterricht gelesener Texte ein eigenes Korpus für Unterrichtszwecke erstellt werden kann. Zudem wird dargelegt, welche Funktionen dieses Programm bietet und wie diese Funktionen für den Sprachunterricht genutzt werden können.

Zum Ende wird erneut die Frage gestellt, ob Korpora nun für die Nutzung im Fremdsprachenunterricht geeignet sind. Es werden Vorteile und Bedenken hinsichtlich der Verwendung von Korpora für den Unterricht einander gegenübergestellt sowie Bedingungen aufgeführt, welche erfüllt sein sollten, damit Korpora sinnvoll in den Unterricht integriert werden können.

1. Korpora und Korpuslinguistik

Bevor eine mögliche didaktische Nutzung von Korpora im Unterricht erörtert werden kann, muss geklärt werden, was unter einem Korpus oder der Korpuslinguistik

im Allgemeinen zu verstehen ist. Daher wird im folgenden Kapitel zunächst auf die Klärung des Begriffs ‚Korpus‘, dessen Verhältnis zur Sprachnorm sowie eine wesentliche Eigenschaft von Korpora, die Repräsentativität, eingegangen werden. Im nächsten Schritt sollen die ‚Korpuslinguistik‘ als Begriff, deren Entwicklung und die Abgrenzung zur theoretischen Linguistik näher durchleuchtet sowie die Frage geklärt werden, ob es sich dabei um eine Methode oder eine Disziplin handelt. Daraufhin sollen die verschiedenen Ansätze der Korpuslinguistik beschrieben und schließlich die Verfahren und Werkzeuge der Korpusaufbereitung und –analyse erklärt werden.

1.1. Was versteht man unter einem Korpus?

Die Arbeit mit Korpora war zunächst auf den englischsprachigen Raum begrenzt, weshalb die grundlegenden Werke über Korpuslinguistik ebenfalls aus dem Englischen stammen. McEnery/Wilson beispielsweise legen in ihrem viel zitierten Werk *Corpus Linguistics* aus dem Jahr 1996 dar, dass empirische Forschung anhand eines beliebigen geschriebenen oder gesprochenen Texts durchgeführt werden kann, wie es zum Beispiel bei der stilistischen Analyse eines einzelnen Gedichts oder eines Romans der Fall ist. Die Vorstellung von einem Korpus als Basis für die empirische Linguistik unterscheide sich allerdings fundamental von der Untersuchung einzelner Texte, so die Autoren. Sie geben daraufhin zunächst eine recht weite Definition des Begriffs Korpus an, wenn sie schreiben: „In principle, any collection of more than one text can be called a corpus: the term ‚corpus‘ is simply the Latin for ‚body‘, hence a corpus may be defined as any body of text“. Trotz der Definition eines Korpus als ein beliebiger ‚Textkörper‘ sprechen McEnery/Wilson allerdings von einer Sammlung von ‚mehr als einem Text‘ und nicht nur von einem einzelnen Text. Sie grenzen diese weite Definition jedoch sogleich ein und definieren den Begriff ‚Korpus‘ im Sinne der modernen Linguistik mit den folgenden Schlagworten: „Sampling and representativeness, finite size, machine-readable form, a standard reference.“²⁴

Die ersten beiden Schlagworte - Stichproben und Repräsentativität - sollen darauf verweisen, dass aufgrund der Unmöglichkeit, alle Äußerungen einer Sprache zu Forschungszwecken zu analysieren, eine Auswahl getroffen werden muss, welche für den jeweiligen Sprachausschnitt repräsentativ ist (zum Begriff der Repräsentativität siehe

²⁴ McEnery/ Wilson (1996): op. cit., S. 21.

Kapitel 1.1.2.). Der nächste Punkt, die begrenzte Größe, spielt auf den Unterschied zwischen statischen Korpora und Monitorkorpora an, welche in Kapitel 2.7. genauer beschrieben werden. Während statische Korpora, wie von McEnery/Wilson gefordert, eine zuvor festgelegte, begrenzte Anzahl an Lexemen umfassen, werden Monitorkorpora laufend erweitert und sind somit, trotz ihrer Aktualität und ihres Nutzens für den Bereich der Lexikographie, keine genaue Quelle in Bezug auf quantitative Analysen. Zudem gehen die Autoren davon aus, dass die heutzutage verwendeten Korpora maschinenlesbar sind, da diese im Vergleich zu gedruckten Korpora zum einen um ein Vielfaches schneller durchsucht und zum anderen schnell und einfach mit zusätzlichen Informationen, wie beispielsweise Annotationen, versehen werden können. Zu guter Letzt geben sie an, dass ein Korpus, auch wenn dies kein essentieller Teil der Definition ist, eine Standardreferenz für die Varietät sein sollte, die es repräsentiert, und somit sollte es möglich sein, dass es von vielen Sprachwissenschaftlern genutzt werden kann und deren Ergebnisse dadurch auch miteinander verglichen werden können.²⁵

Im Jahr 1992 veröffentlichten Geoffrey Leech und Steven Fligelstone ihren Aufsatz „Computers and Corpus Analysis“, in welchem sie Computerkorpora wie folgt definierten:

„Computer corpora are, essentially, bodies of natural language material (whole texts, samples from texts, or sometimes just unconnected sentences), which are stored in machine-readable form. [...] It should be added that computer corpora are rarely haphazard collections of textual material: they are generally assembled with particular purposes in mind, and are often assumed to be (informally speaking) **representative** of some language or text type.“²⁶

Computerkorpora bestehen also aus Texten, Textteilen oder Sätzen natürlicher Sprache, sind maschinenlesbar und gut geplant, sowohl in Bezug auf den Zweck des Korpus als auch in Bezug darauf, welchen Sprachausschnitt oder Texttyp es repräsentieren soll. Zudem geben die Autoren an, dass die Zeilen oder Sätze von Computerkorpora für gewöhnlich nummeriert sein sollten und ein Korpus zusätzlich Informationen grammatischer Natur über Lexeme und Phrasen enthalten kann.²⁷ Die Eckpunkte ihrer Definition sind also ähnlich derjenigen von McEnery/Wilson. Sie führen jedoch genauer aus, dass es sich nicht nur um ganze Texte, sondern auch um

²⁵ Vgl. McEnery/ Wilson (1996): op. cit., S. 21-24.

²⁶ Leech, Geoffrey/ Fligelstone, Steven (1992): „Computers and Corpus Analysis“. In: Butler, C. (Hrsg.) (1992): *Computers and Written Texts*. Oxford: Blackwell, S. 115-116.

²⁷ Vgl. Leech/ Fligelstone (1992): op. cit., S. 116.

Textausschnitte oder bloße Sätze handeln kann.

Für die deutsche Sprache definiert Scherer 14 Jahre später ein Korpus als „eine Sammlung von Texten oder Textteilen, die bewusst nach bestimmten sprachwissenschaftlichen Kriterien ausgewählt und geordnet werden“.²⁸ Hierunter seien nicht nur schriftsprachliche Texte wie Zeitungsartikel, Romane, E-Mails oder Briefe zu verstehen, sondern auch mündliche Äußerungen wie Vorträge, Radiosendungen oder Telefongespräche.²⁹ Bei Scherer wird also besonderes Augenmerk auf die Korpuserstellung gelegt, wenn sie schreibt, dass die Texte nach besonderen Kriterien ausgewählt und geordnet werden. Allerdings schreiben Lemnitzer/Zinsmeister, dass auch ad hoc oder aus anderen sprachwissenschaftlichen Gründen entstandene Textsammlungen als Datenbasis dienen können. Dennoch räumen sie ein, dass diejenigen Korpora, die gut geplant sind, später von größerem Nutzen sind.³⁰

Je mehr zusätzliche Informationen zu den Texten ein Korpus enthält, desto signifikanter ist es für sprachwissenschaftliche Forschungsunternehmen und desto breiter ist das Spektrum der Forschungsvorhaben, für die das Korpus genutzt werden kann. Dabei unterscheidet man einerseits die sogenannten Metadaten wie beispielsweise Informationen über die Autoren der Texte, die Sprecher von Tonaufnahmen, das Datum der Entstehung des geschriebenen oder gesprochenen Texts oder auch die Quelle. Andererseits gibt es sogenannte linguistische Annotationen, also sprachwissenschaftliche Informationen zu Lexemen und Sätzen wie deren linguistische Kategorie oder grammatische Funktion.³¹

Laut Susan Hunston definiert sich ein Korpus durch seine Form und seinen Zweck. Auch für sie ist das elektronisch gespeicherte und zugänglich gemachte Korpus die Norm. Wie bei Scherer und Lemnitzer/Zinsmeister ist ein Korpus ihrer Meinung nach für gewöhnlich geplant, obwohl sie teilweise auch zufällig zustande kommen können, und sie sind zu einem linguistischen Zweck designt worden, der wiederum die Auswahl der Texte bestimmt. Das Ziel ist dabei jedoch nicht nur die Sicherung der Texte, um diese erneut lesen zu können, sondern die Abspeicherung in der Art, dass die Texte nonlinear

²⁸ Scherer (2006): op. cit., S. 3.

²⁹ Vgl. Scherer (2006): op. cit., S. 3.

³⁰ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 8.

³¹ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 8.

gelesen und sowohl quantitativ als auch qualitativ ausgewertet werden können. Andernfalls wäre ein Korpus laut Hunston mit einer Bibliothek oder einem elektronischen Archiv gleichzusetzen.³² In ihrer Erklärung sticht also hervor, dass das Korpus nonlinear gelesen werden kann, im Gegensatz zu Büchern.

Auch Kabatek/Pusch erwähnen die digitale Form in ihrer Definition des Begriffs Korpus:

„Unter einem *sprachlichen Korpus* versteht man in der Linguistik im weiteren Sinne eine (mehr oder minder) systematische Sammlung von sprachlichen Primärdaten; im engeren (und heute dominierenden Sinne) bezieht sich der Begriff auf Sammlungen in *textueller* und *digitaler Form*.“³³

Auch hier ist also wichtig, dass die Daten als Text vorliegen und in der heutigen Zeit digital sein sollten.

Ebenso schreiben Lemnitzer/Zinsmeister, dass die Korpora heutzutage zwar für gewöhnlich in maschinenlesbarer Form vorliegen, es allerdings noch nicht-digitalisierte Textsammlungen gibt. Die Verwendung solcher nicht-digitalisierter Texte führe allerdings zu methodischen Problemen, weshalb viele dieser Texte nachträglich digitalisiert würden. Generell sollte man sich vor einem Forschungsprojekt Gedanken darüber machen, ob digitale Daten für die eigenen Untersuchungen existieren und wenn nicht, wie man diese digitalisieren kann.³⁴ Auch in dieser Arbeit wird das digitale Korpus als die Norm betrachtet, da freier Zugang und uneingeschränkte Verfügbarkeit in Untersuchungen, die sowohl für Lehrer als auch für Schüler nutzbar sein sollen, wichtige Kriterien für die Auswahl der Korpora darstellen.

Lemnitzer/Zinsmeister definieren den Begriff Korpus somit wie folgt:

„Ein Korpus ist eine Sammlung schriftlicher oder gesprochener Äußerungen. Die Daten des Korpus sind typischerweise digitalisiert, d.h. auf Rechnern gespeichert und maschinenlesbar. Die Bestandteile des Korpus bestehen aus den Daten selber sowie möglicherweise aus Metadaten, die diese Daten beschreiben, und aus linguistischen Annotationen, die diesen Daten zugeordnet sind.“³⁵

Es handelt sich hierbei um eine recht weite Definition, da die einzige verpflichtende Richtlinie die ‚Sammlung schriftlicher oder gesprochener Äußerungen‘ darstellt. Die Existenz von Metadaten und Annotationen ist für die Autoren nicht obligatorisch. Auch

³² Vgl. Hunston, Susan (2002): *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press, S. 2.

³³ Kabatek, Johannes/ Pusch, Claus D. (2011): *Spanische Sprachwissenschaft*. Tübingen: Narr Francke Attempto, S. 208.

³⁴ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 8.

³⁵ Lemnitzer/ Zinsmeister (2010): op. cit., S. 8.

1. Korpora und Korpuslinguistik

Stefanowitsch gibt an, dass sprachwissenschaftliche Korpora häufig Annotationen enthalten, welche sich zum einen auf die Sprache selbst beziehen, wie die Angabe der Wortart und Bedeutung jedes Lexems oder die Repräsentation der grammatikalischen Struktur der vorkommenden Sätze, und zum anderen auf außersprachliche Eigenschaften, die sich auf den Text oder auf den Sprecher beziehungsweise Schreiber beziehen, wie Korrekturen und Layouteigenschaften oder Geschlecht, Alter, Bildungsstufe und wirtschaftlicher Status.³⁶ Doch auch seiner Meinung nach können diese zusätzlichen Angaben vorhanden sein, es besteht jedoch für ihn keine Notwendigkeit dazu.

Manuel Alcántara Plá hingegen definiert ein Korpus als „una colección de textos transcritos y anotados con etiquetas a partir de las cuales se extrae información lingüística“.³⁷ Für ihn sind die Annotationen eines Korpus also verpflichtend und Korpora, die lediglich aus Rohdaten bestehen, würden demnach durch das Raster fallen.

Die signifikantesten Charakteristika von Korpora sind laut Scherer deren Größe, Inhalt, Beständigkeit sowie deren Repräsentativität.³⁸ Ein Korpus muss repräsentativ für die darin dargestellte Sprache in ihrer Gesamtheit oder den Ausschnitt einer Sprache sein. So soll ein Korpus der spanischen Standardsprache das Spanische in seiner Gesamtheit entsprechend repräsentieren, während anhand eines spanischen Zeitungskorpus der 90er Jahre lediglich allgemeingültige Aussagen über diesen Sprachausschnitt getroffen werden sollen.

Auch Elena Tognini-Bonelli beschreibt ein Korpus als eine Sammlung von Texten, von denen angenommen wird, dass sie für eine gegebene Sprache repräsentativ sind, sodass sie für sprachwissenschaftliche Untersuchungen verwendet werden können. Normalerweise handelt es sich dabei um natürlich vorkommende Sprachausschnitte, die nach genauen Designkriterien gesammelt wurden, zu einem bestimmten Zweck und unter der Annahme, dass sie größere Sprachausschnitte repräsentieren. Dabei wird der

³⁶ Vgl. Stefanowitsch, Anatol (2005): „Quantitative Korpuslinguistik und sprachliche Wirklichkeit“. In Solte-Gresser, C./ Struve, K./ Ueckmann, N. (Hrsg.) (2005): *Von der Wirklichkeit zur Wissenschaft. Aktuelle Forschungsmethoden in den Sprach-, Literatur- und Kulturwissenschaften*. Münster: LIT Verlag, S. 148.

³⁷ Alcántara Plá, Manuel (2007): *Introducción al análisis de estructuras lingüísticas en corpus. Aproximación semántica*. Madrid: UAM, S. 6.

³⁸ Vgl. Scherer (2006): op. cit., S. 5-6.

Anspruch erhoben, dass diese Sammlung „authentische“³⁹ Sprache repräsentiert.⁴⁰ Auch sie nimmt in ihre Definition also den Begriff der Repräsentativität der Daten auf. Zusätzlich ist für sie der Begriff der Authentizität des Sprachmaterials von besonderer Bedeutung, welcher in Kapitel 3.2.1. noch genauer betrachtet werden soll.

Für Stefanowitsch zeichnen sich sprachwissenschaftliche Korpora durch zwei besondere Eigenschaften aus, welche mit der Repräsentativität der Korpora zusammenhängen: Zum einen enthalten sie, wie bei Tognini-Bonelli, nur Texte, die in natürlichen Sprachverwendungssituationen entstanden, also nicht eigens zur Erstellung des Korpus erzeugt worden sind, und zum anderen wird bei diesen Korpora ein realistischer Querschnitt über verschiedene sprachliche Genres, Themen, Register und Modalitäten hinweg angestrebt. Allerdings räumt er ein, dass zum Zweck mancher Untersuchungen das zweite Kriterium vernachlässigt werden muss, dann zum Beispiel, wenn gezielt nach der Verwendung von Strukturen in bestimmten Genres oder Registern gesucht werden soll und diese Genres oder Register somit bevorzugt oder ausschließlich in das Korpus integriert werden. Das erste Kriterium ist allerdings grundlegend und zeichnet die Korpuslinguistik an sich aus, weshalb Abweichungen von dieser Richtlinie selten sind.⁴¹

Hunston legt ebenfalls Wert auf die Natürlichkeit der Entstehung von Texten, wenn sie schreibt, Linguisten hätten schon immer das Lexem ‚Korpus‘ verwendet, um eine Sammlung natürlich auftretender Sprachbeispiele, angefangen von ein paar Sätzen bis hin zu einer Reihe geschriebener Texte oder Tonbandaufnahmen, zu bezeichnen.⁴² Es muss sich also nicht um eigens zum Zweck der Korpuserstellung geschriebene oder gesprochene Texte handeln. Wie wir bei der Kategorisierung der Korpora sehen werden, gibt es dennoch Korpora, bei denen die Texterstellung zur Entwicklung eines Korpus legitim ist, wie zum Beispiel bei Lernerkorpora.

Für die vorliegende Arbeit wird ein Korpus nun als digitale Sammlung von schriftlichen oder mündlichen Texten definiert, welche für eine bestimmte Sprache oder

³⁹ Der Begriff der Authentizität wird in Unterkapitel 3.2.1. näher beleuchtet, es soll jedoch bereits an dieser Stelle darauf hingewiesen werden, dass anstatt der umstrittenen Bezeichnung ‚authentisch‘ der Begriff ‚real‘ bevorzugt und im Folgenden verwendet wird.

⁴⁰ Vgl. Tognini-Bonelli, Elena (2001): *Corpus Linguistics at Work*. Amsterdam/ Philadelphia: John Benjamins, S. 2.

⁴¹ Vgl. Stefanowitsch (2005): op. cit., S. 148.

⁴² Vgl. Hunston (2002): op. cit., S. 2.

eine sprachliche Varietät repräsentativ ist. Dabei kann ein Korpus zusätzlich zu den Texten in Rohform über Metadaten und/oder linguistische Annotationen verfügen.

In den gängigen Definitionen des Begriffs ‚Korpus‘ sowie in der für diese Arbeit gewählten Begriffserklärung spielt die Repräsentativität desselben also eine wichtige Rolle. Somit muss sich der Nutzer von Korpora erst einmal die Frage stellen, was Repräsentativität ist und woran man sie misst. Im Wörterbuch für Erziehung und Unterricht wird folgende Definition für den Begriff ‚Repräsentativität‘ angegeben: „Repräsentativ sind solche Ergebnisse, Gegenstände oder Gruppen, die für alle anderen Personen oder Sachverhalte der Bereiche, in denen Untersuchungen angestellt werden, stellvertretend stehen.“⁴³ Für Korpora heißt das, wie bereits bei der Definition des Korpus erwähnt wurde, dass sie repräsentativ sein sollen für die jeweilige Sprache oder die Varietät, die sie darstellen. Um sich allerdings mit der Repräsentativität von Korpora beschäftigen zu können, muss zunächst die Frage nach der Repräsentativität sprachlicher Daten im Allgemeinen und damit nach der Sprachnorm und dem Verhältnis derselben zu den in Korpora präsentierten Sprachdaten geklärt werden.

1.1.1. Korpora und Sprachnorm

Hundt (2008) betrachtet Grammatikalität, Akzeptabilität und Sprachnorm im Hinblick auf das Zusammenspiel zwischen der vorwiegend theorieorientierten Linguistik und der empiriebasierten Korpuslinguistik (für eine genaue Unterscheidung beider Arten der Linguistik siehe Kapitel 1.2.1. und 1.2.2.). Seiner Meinung nach ist die Zusammenarbeit zwischen Linguisten beider Lager besonders an den Rändern der Sprachnorm gefragt. Dieser Bereich wird zwar noch von der *langue* abgedeckt, fällt jedoch nicht mehr oder noch nicht in die Sprachnorm. Hierfür untersucht Hundt zunächst die drei Begriffe Grammatikalität, Akzeptabilität und Sprachnorm.⁴⁴ Generell seien Grammatikalitätsurteile varietätenbezogen, individuell, graduierbar, wandelbar und hoch streuend, was vor allem daran festzumachen sei, dass jeder Sprecher eine andere Meinung zur Grammatikalität bestimmter Äußerungen habe. Zudem stelle sich die Frage nach der Grammatikalität besonders in den Grenzbereichen der Sprachnorm, also bei eben jenen

⁴³ Köck, Peter (2008): *Wörterbuch für Erziehung und Unterricht*. Augsburg: Brigg Pädagogik Verlag, S. 414.

⁴⁴ Vgl. Hundt, Markus (2008): „Grammatikalität – Akzeptabilität – Sprachnorm. Zum Verhältnis von Korpuslinguistik und Grammatikalitätsurteilen“. In: Lenz, F./ Schierholz, S. (Hrsg.) (2008): *Corpuslinguistik in Lexik und Grammatik*. Tübingen: Stauffenburg, S. 15-16.

noch grammatischen Konstruktionen, die nicht mehr oder noch nicht im Bereich der Sprachnorm liegen. Konstruktionen im Zentrum der Sprachnorm, welche sich durch volle Akzeptanz durch die Sprachgemeinschaft und volle Grammatikalisierung auszeichnen, sowie andererseits Konstruktionen, die weit außerhalb der *langue* liegen (also zweifelsfrei ungrammatische Konstruktionen, die vom Autor als „konstruktionelle ‚Eintagsfliegen‘ oder schlichtweg als performanzbedingte Fehler“⁴⁵ bezeichnet werden und der *parole* zuzuschreiben sind), werden von befragten Studierendengruppen laut Hundt eindeutig als grammatisch korrekt im ersten beziehungsweise ungrammatisch im zweiten Fall bewertet. Während unproblematische Konstruktionen innerhalb der Sprachnorm allerdings nicht markiert sind als ‚grammatisch korrekt‘, sondern unmarkiert bleiben, also schlichtweg als grammatisch-korrekte Konstruktionen passieren, sind es die Problemfälle in ihrem zweifelhaften Grammatikalitätsstatus, welche markiert sind, ebenso wie ungrammatische Konstruktionen. Grammatikalität lässt sich nach Hundts Meinung somit am besten an den Grenzen der Grammatikalität erkunden.⁴⁶ Auch laut Busse wird das Verhältnis zwischen Sprachsystem und Sprachnorm vor allem im Verhältnis zwischen dem ‚Standard-‘ oder ‚Normalfall‘ eines linguistischen Phänomens und den sogenannten ‚Ausnahmefällen‘ ersichtlich. Er gibt jedoch zudem an, dass es nicht gerade leicht ist, bei einem konkreten sprachlichen Phänomen zu entscheiden, was der Normalfall ist und was nicht, und dass diese Entscheidung meist auf der Intuition der Sprachwissenschaftler basiert. Dieses von ihm so betitelte „prekäre Verhältnis“ zwischen Normfall und Abweichung ist ein Grundproblem jeder Normbeschreibung, und man muss sich bei jedem Sonderfall fragen, wie weit die Grenzen der Norm gezogen werden und werden können.⁴⁷

In Bezug auf die Sprachnorm stellt sich für Hundt die Frage nach der Akzeptabilität von Konstruktionen, dem Normtypus und der Frequenz. Dabei ist die Akzeptabilität einer Konstruktion der wesentliche Indikator für die Sprachnorm, wobei sich im Bereich Letzterer alle Konstruktionen befinden, die auch für die Mehrheit der Sprecher akzeptabel sind. Die Akzeptabilität von Konstruktionen wird über Probandenbefragungen ermittelt, wobei diese, wie die Grammatikalitätsurteile, großen Schwankungen unterliegen. Es

⁴⁵ Hundt (2008): op. cit., S. 20.

⁴⁶ Vgl. Hundt (2008): op. cit., S. 17-21.

⁴⁷ Busse, Dietrich (2006): „Sprachnorm, Sprachvariation, Sprachwandel. Überlegungen zu einigen Problemen der sprachwissenschaftlichen Beschreibung des Deutschen im Verhältnis zu seinen Erscheinungsformen“. In: Deutsche Sprache, 34. Jg., Heft 4/2006, S. 315, 318.

werden auch Textkorpora herangezogen, da die Frequenz einer Konstruktion ebenso als Indiz für ihre Akzeptabilität gilt. Auch die Akzeptabilität von Konstruktionen unterliegt Veränderungen unter anderem aufgrund von Frequenzveränderungen (was man häufig hört, erscheint akzeptabler als selten Gehörtes), aufgrund des *tacit knowledge* der individuellen internen Grammatik der Sprecher (welches nur bedingt und nicht immer vollständig zugänglich ist) sowie aufgrund metasprachlicher Urteile (beispielsweise in Grammatiken).⁴⁸ In Bezug auf das *tacit knowledge*, also das intuitive Wissen der Sprecher, schreibt Felder, dass die Notwendigkeit von Sprachnormen den meisten Sprachteilnehmern unbewusst gegenwärtig ist.⁴⁹ Sprachnormen werden als Manifestationen des Akzeptablen in Grammatiken und Wörterbüchern schriftlich fixiert, darin wird also das *tacit knowledge* der Sprecher rekonstruiert. In der Regel spiegeln die Fixierungen allerdings nur wider, was bereits als akzeptabel gilt, bisweilen wird die Akzeptabilität einer Konstruktion jedoch auch durch ihre Niederschrift in einer Grammatik gesteigert, ganz gleich, ob es sich dabei um eine deskriptive oder präskriptive Grammatik handelt.⁵⁰ Busse gibt zudem an, dass im Vergleich zur existierenden Norm einer Standardsprache, wie sie beispielsweise in der Schule vermittelt wird, die verfügbaren Kodifikationen in Form von Grammatiken, Wörterbüchern oder Lehrwerken nur jeweils einen Ausschnitt darstellen. Zum einen ist die Gesamtheit der Normen zu keinem Zeitpunkt vollständig zugänglich, zum anderen widersprechen sich die verfügbaren Kodifikationen teilweise, wodurch kein einheitliches Bild entstehen kann.⁵¹

Der Normtypus, Hundts zweiter die Sprachnorm beeinflussender Faktor, ist von Bedeutung für die Sprachnorm in der Schnittmenge zwischen System und Rede. Hundt bezieht sich in seinen Ausführungen auf das von Eroms (2000) aufgestellte Modell von System, Norm und Rede, welcher wiederum das Modell von Coseriu (1974) aufgreift.⁵² Coseriu, auf den die Erweiterung der Dichotomie zwischen *langue* und *parole* zurückgeht, befand das Saussuresche Modell als ungenügend und ergänzte daraufhin die Norm als Schnittmenge zwischen System und Rede. Als Beispiel führt er die offene oder

⁴⁸ Vgl. Hundt (2008): op. cit., S. 21-22.

⁴⁹ Vgl. Felder, Ekkehard (2003): „Das Spannungsverhältnis zwischen Sprachnorm und Sprachvariation als Beitrag zu Sprach(differenz)bewusstheit“. In: Wirkendes Wort, 53. Jg., Heft 3/2003, S. 494.

⁵⁰ Vgl. Hundt (2008): op. cit., S. 22.

⁵¹ Vgl. Busse (2006): op. cit., S. 319.

⁵² Vgl. Eroms, Hans-Werner (2000): *Syntax der deutschen Sprache*. Berlin: De Gruyter. Und: Coseriu, Eugenio (1974): *Synchronie, Diachronie und Geschichte: Das Problem des Sprachwandels*. München: Fink.

geschlossene Realisierung der Vokale ‚e‘ und ‚o‘ im Spanischen an. Während das ‚o‘ geschlossen ist in *llamó*, *boda* und *esposa*, ist es in *rosa*, *hoja* und *dogma* offen. Es gibt im Spanischen allerdings keine funktionelle Opposition zwischen offenen und geschlossenen Vokalen. Würden *llamó*, *boda* oder *esposa* offen ausgesprochen, so würde das System nicht verletzt, da die Lexeme ihre Bedeutung nicht verändern würden und man sie auch noch erkennen könnte. Es würde sich lediglich um eine „nicht-normale Realisierung“ handeln. Es gibt im phonologischen System des Spanischen somit nur ein einziges Phonem /o/, das in Opposition steht zu /a/, /e/, /i/ und /u/ (so wird beispielsweise *ojo* von *ajo* oder *oso* von *uso* unterschieden). Allerdings gibt es zwei konstante Invarianten in der Norm, die funktionell zu a, e, i, u und untereinander auch nicht-funktionell in Opposition stehen. Auf der Ebene der Rede, wird wiederum eine unendliche Anzahl an verschiedenen offenen und geschlossenen Varianten des Vokals ‚o‘ realisiert, welche jedoch auf der Ebene der Norm auf einen einzigen allgemeinen Typ des offenen oder geschlossenen ‚o‘ zurückgeführt wird.⁵³ Bei der Erstellung seines Modells stützt sich Eroms (2000) auf die folgenden Überlegungen Coserius: Eine Einzelsprache ist ein „Diasystem“ mit synchronen, diachronen, diatopischen und funktionalen Dimensionen. Es kann über zwei Vorkommensbereiche erschlossen werden, zum einen über die Beobachtung, zum anderen über die Bewertung. Was beobachtet oder auch selbst verwendet wird, ist die Rede (*parole*). Was als „üblich und gut“ bewertet wird, ist die Norm. Das System ist schließlich, „was synchron verbindlich, diachron belegbar und regional, fach- und individualspezifisch möglich ist“.⁵⁴

Während bei Coseriu allerdings einerseits das System über die Norm hinausgeht, da es nicht nur die traditionellen Realisierungen enthält, sondern auch solche, die nicht oder noch nicht traditionell sind (aufgrund der Tatsache, dass es ein „System von Möglichkeiten“ darstellt), sondern auch die Norm über das System hinausgeht, wie im Fall der Zusatzinformation über die offene oder geschlossene Aussprache des Vokals ‚o‘ im obigen Beispiel, stellt die Norm bei Eroms (2000) und damit auch bei Hundt (2008) eine Schnittmenge aus System und Rede dar, und geht somit nicht über das System

⁵³ Coseriu, Eugenio (1970): *Sprache – Strukturen und Funktionen*. Tübingen: Narr, S. 193-212.

⁵⁴ Eroms (2000): op. cit., S. 18.

hinaus.⁵⁵ Eroms (2000) und Hundt (2008) stellen ihr Modell von System, Norm und Rede wie folgt dar:

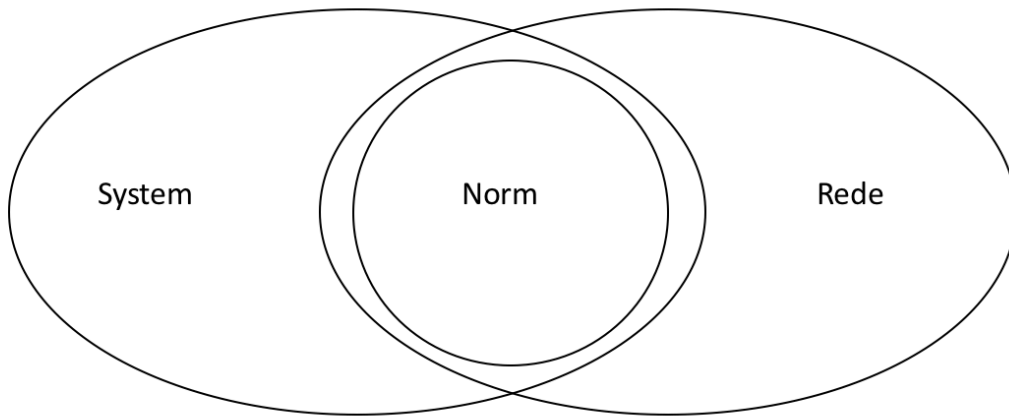


Abbildung 1: System – Norm – Rede nach Eroms (2000).⁵⁶

Während das Sprachsystem alle möglichen, systemkonformen Konstruktionen umfasst, und unter dem Begriff ‚Rede‘ alle realisierten Äußerungen, also das konkrete Sprechen, zusammengefasst werden, stellt die Sprachnorm in diesem Modell eine Schnittmenge aus System und Rede dar, da sie systemkonforme und akzeptable Konstruktionen in sich vereint. Die hierin liegenden Äußerungen, müssen nicht nur das Kriterium der Systemkonformität erfüllen, sondern sie müssen zudem tatsächlich realisiert werden, also Teil der *parole* sein, was an ihrer Frequenz erkannt werden kann.⁵⁷ Im Hinblick auf das Hinausgehen der Norm über das System muss allerdings hinzugefügt werden, dass es sich in diesem Modell laut Hundt um ein „Kontinuum zwischen Sprachsystem, Sprachnorm und Rede“ handelt.⁵⁸ Des Weiteren schreibt Hundt im Jahr 2010, dass diejenigen Konstruktionen, welche sich nicht in der Schnittmenge zwischen System und Rede befinden, auch nicht Teil der Norm sind, da sie nicht wohlgeformt sind und daher nicht vom Sprachsystem abgedeckt werden.⁵⁹ Er bezieht sich mit der genauen Abbildung der Norm innerhalb des Systems somit vor allem auf die Wohlgeformtheit der Konstruktionen.

⁵⁵ Forschungsstelle Eugenio Coseriu (ed.2016): *System, Norm und Rede*. Online Publikation: <http://www.coseriu.de> (12.07.2018).

⁵⁶ Eroms (2000): op. cit., S. 18.

⁵⁷ Vgl. Hundt (2008): op. cit., S. 16.

⁵⁸ Hundt (2008): op. cit., S. 39.

⁵⁹ Vgl. Hundt, Markus (2010): „New norms – How new grammatical constructions emerge“. In: Lenz, A. N./ Plewnia, A. (Hrsg.) (2010): *Grammar between Norm and Variation*. Frankfurt: Lang, S. 32.

Im Hinblick auf den Normtypus gilt es laut Hundt, gesetzte von konventionellen Normen zu unterscheiden. In den Fachterminologien, insbesondere im Bereich der Fachlexik, handelt es sich häufig um gesetzte, von Gremien festgelegte Normen. Für die Entwicklung der Grammatik, beispielsweise der deutschen Standardsprache, sind jedoch meist konventionelle Normen ausschlaggebend, das heißt, diese haben sich aus dem Sprachgebrauch heraus ergeben. Somit steht fest, dass obgleich die Sprachteilnehmer die in Grammatiken niedergeschriebenen Normen als gesetzte Normen auffassen, die Sprachnormen der Grammatik der Standardsprache doch einen konventionellen Ausgangspunkt haben. Wären die Normen syntaktischer Konstruktionen gesetzt, so gäbe es keinen Grenzbereich der Sprachnorm, und es wäre in jedem Fall eindeutig, was innerhalb und was außerhalb der Sprachnorm liegt. Der Charakter der Sprachnormen in der Alltagssprache ist also in der Regel konventionell. Dies variiert allerdings von Land zu Land. Aussagen von Sprachexperten oder auf irgendeine Weise legitimierten Gremien können der Realität hier nicht gerecht werden.⁶⁰ Auch Busse gibt an, dass die Sprachnorm der Sprachvariation gegenüber nicht neutral ist, da die Frage nach der möglichen und zulässigen Variation die Beschreibung der Norm ständig berührt aufgrund der Tatsache, dass jedes Grammatikalitätsurteil implizit zur Abgrenzung der Norm von anderen Varianten beiträgt.⁶¹ Zurückliegende Sprachvariationen können heute in der Sprachnorm enthalten sein, so Felder, der das Spannungsverhältnis zwischen Sprachnorm und Sprachvariation grundsätzlich als ein fruchtbares beschreibt, da man Sprachnormen „auf kreative und bewusste Weise und mit erkennbarer Sprecherintention durchbrechen bzw. modifizieren oder variieren kann“.⁶²

Schließlich trägt Hundts dritter Faktor, die Frequenz von Konstruktionen, maßgeblich zur Akzeptanzsteigerung bei, speziell bei neuen Konstruktionstypen. Als Beispiel führt er das Dativ-/Rezipientenpassiv im Deutschen an, das einen sogenannten Frequenzdurchbruch erfahren hat und nun zumindest so frequent ist, dass das Problem der geringen Belegdichte nicht mehr angeführt werden kann, obwohl diese syntaktische Konstruktion auch heute noch nicht 1000fach belegt ist. Allerdings ist die Frequenz als Indikator für die Sprachnorm nicht unproblematisch, da sich keine Grenze oder Anzahl von Vorkommen festlegen lässt, ab der man eine Konstruktion der Sprachnorm zuordnen

⁶⁰ Vgl. Hundt (2008): op. cit., S. 22-23.

⁶¹ Vgl. Busse (2006): op. cit., S. 319.

⁶² Felder (2003): op. cit., S. 494-495.

kann. Die Frage, wie oft sie in einem Korpus vorkommen muss, um die Sprachnormhaftigkeit festzustellen, kann nicht endgültig geklärt werden, doch auch Frequenzanstiege in bestimmten Intervallen, wie im Fall des Dativ-/Rezipientenpassivs, können als Indiz herangezogen werden. Für die Grammatikalität einer Konstruktion spielt die Frequenz hingegen keine Rolle, da hier die zugrunde liegende syntaktische Regularität ausschlaggebend ist.⁶³

Hundt geht von einem weiten Sprachnormbegriff aus, in welchem auch Nichtstandardvarietäten genormt sein können, sofern sie über ein Kriterium der Akzeptabilität normkonforme von normabweichenden Konstruktionen unterscheiden. Es fallen solche Konstruktionen in die Sprachnorm, die grammatisch und akzeptabel sind. Indizien für die Akzeptabilität sind Gebrauchs-Frequenz, Bildungsregularität und Ausgrenzung ungrammatischer Konstruktionen.⁶⁴ Felder merkt zudem an, dass die Sprachnorm nicht als unveränderliche Konstante aufzufassen ist, sondern als dynamische Größe, die stets zur Verfügung steht und ihre Berechtigung immer wieder aufs Neue unter Beweis stellt.⁶⁵ Diese Veränderlichkeit der Sprachnorm wurde bereits von Coseriu (1970) konstatiert.⁶⁶

Hundt geht außerdem auf die Verwendbarkeit von Korpora zur Untersuchung grammatischer Grenzphänomene ein. Für ihn sind für diesen Zweck lediglich maschinenlesbare, „authentische“ Textkorpora geeignet (zum Begriff der Authentizität siehe auch 3.2.1.), welche er von elizitierten Korpora unterscheidet. Letztere bezeichnet er als „selbst zusammengestellte Belegreihen“, da er diese eigens zum Zweck des jeweiligen Forschungsvorhabens zusammengestellten Textsammlungen nicht als Korpora im eigentlichen Sinn versteht.⁶⁷ Elizitierte Korpora könnten seiner Meinung nach durchaus Aufschluss über Fragen der Akzeptabilität geben, jedoch könne daran nicht der tatsächliche Sprachgebrauch überprüft werden. Allerdings seien auch bei Akzeptabilitätsurteilen „authentische“ Korpora den elizitierten vorzuziehen, sofern sie zur Verfügung stünden. Auf dieser Basis stellt Hundt drei Kriterien zur Nutzung von Korpora im Hinblick auf die Sprachnorm auf. Erstens ist ein Korpus ein Korrektiv, das heißt, es bewahrt den Nutzer zum einen vor Phantasiekonstruktionen, welche zwar

⁶³ Vgl. Hundt (2008): op. cit., S. 23-24.

⁶⁴ Vgl. Hundt (2008): op. cit., S. 24-25.

⁶⁵ Vgl. Felder (2003): op. cit., S. 475.

⁶⁶ Vgl. Coseriu (1970): op. cit., S. 193-212.

⁶⁷ Hundt (2008): op. cit., S. 27.

grammatisch denkbar, aber inakzeptabel sind, und zum anderen kann es Belege liefern für Konstruktionen, die zwar nicht einem grammatischen Modell entsprechen, aber aufgrund ihrer Frequenz ernst zu nehmen sind. Zweitens sagen Korpora nicht, was grammatisch und was ungrammatisch ist, denn die Tatsache, dass eine Konstruktion nicht im Korpus erscheint, heißt nicht, dass sie in der jeweiligen Varietät nicht grammatisch ist.⁶⁸ In diesem Zusammenhang warnt Johns (1991) im Hinblick auf die Analyse und Interpretation von Korpusdaten mit Schülern vor Übergeneralisierungen, da das Fehlen bestimmter Strukturen in einem Korpus nicht mit deren Inexistenz und Ungrammatikalität gleichzusetzen ist. Vielmehr ist in diesem Zusammenhang von ‚Normalität‘ zu sprechen. Häufig im Korpus erscheinende Strukturen sind demnach die ‚normale‘, also gebräuchliche Art und Weise der Versprachlichung eines Sachverhalts, während die Regeln einer Sprache durchaus andere Möglichkeiten zulassen würden.⁶⁹ Zudem enthalten Korpora, insbesondere solche mündlicher Art, auch ungrammatische Konstruktionen, welche die Untersuchungsergebnisse jedoch aufgrund ihrer singulären Erscheinung nicht negativ beeinflussen. Drittens ist die von Hundt sogenannte Korpuschichtung von Bedeutung, da nur dann von der Grammatikalität einer Konstruktion gesprochen werden kann, wenn sie zumindest in zeitlich, wenn nicht auch in sozial und funktional geschichteten Korpora festgestellt werden kann.⁷⁰ Busse merkt in diesem Kontext an, dass die Sprachnorm im Sinne der heutigen Grammatik, ein „historisches Entwicklungsprodukt“ ist, welches sich durch die Ausgrenzung früher üblicher grammatischer Möglichkeiten herausgebildet hat.⁷¹

Hundt untersucht auf der soeben beschriebenen Basis deutsche Konstruktionen wie das Reflexivpassiv, das Dativpassiv oder Wortartenübergänge und klärt für jede Kategorie den gegenwärtigen Status der Grammatikalität. Im Grenzbereich der Grammatikalität spielen also die Grammatikalisierung der entsprechenden Konstruktion, die Frequenz derselben sowie die Akzeptanz durch die Sprachteilnehmer zusammen und verorten die Konstruktion im Übergangsbereich zwischen grammatischen und nicht-akzeptablen Konstruktionen einerseits und grammatischen und akzeptablen

⁶⁸ Vgl. Hundt (2008): op. cit., S. 26-27.

⁶⁹ Vgl. Johns, Tim (1991): „Should you be persuaded?: Two samples of data-driven learning materials“. In: Johns, T./ King, P. (Hrsg.) (1991): *Classroom Concordancing ELR Journal*, 4. University of Birmingham, S. 1-16.

⁷⁰ Vgl. Hundt (2008): op. cit., S. 27-28.

⁷¹ Busse (2006): op. cit., S. 319.

Konstruktionen andererseits. Für Hundt steht fest, dass zu Beginn dieses Prozesses eine nicht genauer spezifizierte zugrunde liegende syntaktische Regularität steht. Je mehr die Konstruktion akzeptiert wird, desto weiter schreitet der Grammatikalisierungsprozess voran und schlägt sich schließlich in der Anzahl der Belege in Textkorpora nieder. Somit kann erst am Ende des Prozesses anhand der Korpora festgestellt werden, dass sich die Konstruktion zu einer grammatischen und akzeptablen gewandelt hat.⁷²

So kommt Hundt zu dem Schluss, dass introspektiv und empirisch arbeitende Linguisten einander insbesondere im Grenzgebiet zwischen Grammatikalität und Akzeptanz ergänzen können. Während erstere Auskunft über syntaktische Regularitäten und die Ablehnung von Konstruktionen geben können, liefern Korpuslinguisten die Frequenzen strittiger Phänomene und geben an, über welche Konstruktionen es sich nachzuforschen lohnt.⁷³

Auch Heringer (2009) plädiert für die Nutzung von Korpora als empirische Basis grammatikalischer Untersuchungen. Er veranschaulicht diese Notwendigkeit anhand der exemplarischen Ermittlung der korrekten Genitivform für den Ausdruck *Walther von der Vogelweide* in aktuellen Grammatiken. Dabei macht er darauf aufmerksam, dass die Grammatikschreibung unvollständig ist, die Sprachwissenschaftler sich dessen vollkommen bewusst sind und dennoch nichts dagegen unternehmen, da es sich bei einer Grammatik lediglich um ein theoretisches Konstrukt handelt, welches durch praktische Zwecke bestimmt ist. Obwohl er nach eigenen Aussagen kein Korpuslinguist ist, vertritt er die Meinung, dass die interaktive Untersuchung von Korpora lehrreich ist. Seiner Ansicht nach sollten Korpusdaten die Quelle aller sprachwissenschaftlichen Untersuchungen sein. Es sollte alle Information aus den Korpora abgeleitet werden, ohne zu viel an linguistischem Wissen hineinzustecken, um so mittels einer automatisierten Anwendung des Wissens zu Ergebnissen zu kommen. Auch wenn es sich hierbei lediglich um ein Ideal handelt, so sollten Algorithmen entwickelt werden, die aus den Korpusdaten Ergebnisse ableiten, wobei man in dieser Hinsicht in Bezug auf die Grammatik noch nicht weit fortgeschritten ist. Zum Ende seines Aufsatzes ruft er den Leser dazu auf, „den

⁷² Vgl. Hundt (2008): op. cit., S. 28-29.

⁷³ Vgl. Hundt (2008): op. cit., S. 39.

langen Marsch durch die Intuitionen“ zu beenden und Korpora intelligent zu nutzen – auch für die Grammatik.⁷⁴

Kupietz/Keibel sind ebenso der Meinung, man müsse sich von der Empirie leiten lassen, um zu einer explanatorischen Theoriebildung zu gelangen und sich mit möglichst wenigen Vorannahmen dem Untersuchungsgegenstand Sprache zu nähern. Nur auf diese Weise könne die Kluft zwischen der angestrebten theoretischen Beschreibung und der Ebene des Sprachgebrauchs so klein wie möglich gehalten werden. Sie gehen davon aus, dass Sprache als System im Sinne der *langue* weder statisch ist noch unabhängig von und zeitlich vor dem Sprachgebrauch existiert. Somit gibt es keine Möglichkeit, das Phänomen Sprache unabhängig von seinem Gebrauch zu definieren. Da das Sprachsystem nicht unmittelbar erfassbar ist, sondern nur mittelbar über den konkreten Sprachgebrauch, ist Letzterer nicht nur die Realisierungsmöglichkeit der *langue*, sondern gleichzeitig ihre primäre Ursache. Demnach wirkt sich die individuelle Spracherfahrung jedes Sprachteilnehmers auf dessen prozedurales Sprachwissen aus, wodurch jeder Sprecher sozusagen eine individuelle Grammatik in seinem Kopf hat, die wiederum seine Sprachproduktion beeinflusst und zur Spracherfahrung anderer Sprecher beiträgt. Das Sprachsystem an sich zeigt sich in Form von Konventionen der Sprachgemeinschaft, welche prinzipiell in jeder sprachlichen Interaktion ausgehandelt werden. In diesem Sinne gibt es laut Kupietz/Keibel keine grammatischen „Regeln“ als stabile verbindliche Gebilde, sondern stattdessen „Regelhaftes“ im Sprachgebrauch. Diese strukturellen Konventionen, bei denen der Grad der Konventionalisierung stark variieren kann, sind ihrer Meinung nach grundsätzlich instabil, dynamisch, kontextabhängig und adaptiv.⁷⁵

Kupietz/Keibel bezeichnen als „Sprache“ (in Anführungszeichen) sowohl den beobachtbaren Sprachgebrauch, vergleichbar mit Coserius individueller Ebene der konkreten Redeereignisse und damit der einzigen „wirklich konkret sich manifestierenden Form von Sprache“ gemäß Kabatek/Pusch, als auch das emergente Sprachsystem, im Sinne von Coserius historischer Ebene der Einzelsprachen.⁷⁶ Um nun

⁷⁴ Heringer, Hans Jürgen (2009): „Ist das Deutsche grammatisch zu erfassen?“. In: Konopka, M./ Strecker, B. (Hrsg.) (2009): *Deutsche Grammatik – Regeln, Normen, Sprachgebrauch*. Berlin: De Gruyter, S. 31.

⁷⁵ Kupietz, Marc/ Keibel, Holger (2009): „Gebrauchsbasierte Grammatik: Statistische Regelhaftigkeit“. In: Konopka, M./ Strecker, B. (Hrsg.) (2009): *Deutsche Grammatik – Regeln, Normen, Sprachgebrauch*. Berlin: De Gruyter, S. 33-41.

⁷⁶ Kabatek/ Pusch (2011): op. cit., S. 32.

über die ‚Sprache‘ etwas aussagen zu können, muss laut Kupietz/Keibel auf einer unteren Ebene ein passendes Korpus gewählt werden, anhand dessen Hypothesen überprüft werden können. Sie sind darüber hinaus der Ansicht, dass für eine explanatorische Theoriebildung auch eine darüberliegende, psychologische Ebene mit einbezogen werden muss, um die Gefahr einer blinden Induktion zu vermeiden. Für das Forschungsprogramm von Kupietz/Keibel, mittels Beobachtungen auf der Korpusebene induktiv zu Hypothesen auf der Ebene der ‚Sprache‘ zu gelangen, soll das Korpus stellvertretend für die Grundgesamtheit des relevanten Ausschnitts des Sprachgebrauchs stehen. Es soll ermittelt werden, zu welchen Generalisierungen Sprachteilnehmer gelangen würden, wenn sich ihre Spracherfahrung aus den Korpusdaten speisen würde, und dieses Konzept des abstrahierten Sprachteilnehmers soll für die einzelnen Mitglieder der Sprachgemeinschaft stehen. Sie beschreiben für ihr Forschungsprogramm psychologische Prämissen, wie die Fähigkeit, Ähnlichkeiten zu erkennen, und bezeichnen die Generalisierungen über diese Ähnlichkeitsbeziehungen als Präferenzrelationen. Diese Präferenzrelationen vereinen die Konzepte von Regeln, Normen und deren Varianten in sich und beinhalten darüber hinaus weniger auffällige Realisierungen, die meist der linguistischen Beschreibung entgehen. Während Regeln und Normen jedoch harte, kategorielle Relationen darstellen, sind Präferenzrelationen grundsätzlich graduell und kontextabhängig. Kupietz/Keibel zufolge besteht alles Regelhafte in einer Sprache aus Präferenzrelationen. Die Autoren versuchen also mittels einer empirischen Herangehensweise in kleinen Schritten zu einer explanatorischen Theorie von ‚Sprache‘ und insbesondere von Grammatik zu gelangen, wobei sie sich sicher sind, dass diese Vorgehensweise in den kommenden Jahren zunehmend vereinfacht wird durch die Fortschritte im Bereich der Korpuslinguistik.⁷⁷

Wie vor allem an den Ausführungen von Hundt (2008) und Kupietz/Keibel (2009) zu sehen war, können Korpora Auskunft über die Akzeptabilität von Konstruktionen geben und somit auch Rückschlüsse auf die Sprachnorm liefern. Die geringe Anzahl inkorrektur Belege kann laut Hundt vernachlässigt werden⁷⁸, wonach Korpora im Hinblick auf die Verortung von Konstruktionen auf der Skala der Akzeptabilität nicht zu unterschätzen sind und damit durchaus für den schulischen Kontext genutzt werden

⁷⁷ Vgl. Kupietz/ Keibel (2009): op. cit., S. 42-50.

⁷⁸ Vgl. Hundt (2008): op. cit., S. 27.

können. Nachdem nun näher auf die Sprachnorm und das Verhältnis der Korpora zur selben eingegangen wurde, kann auf die eingangs gestellte Frage nach der Repräsentativität von Korpora zurückgekommen werden.

1.1.2. Repräsentativität von Korpora – zwingende Eigenschaft oder unerreichbares Ziel?

Rieger zitiert in seinem Aufsatz *Repräsentativität: von der Unangemessenheit eines Begriffs zur Kennzeichnung eines Problems linguistischer Korpusbildung* aus dem Jahr 1979 eine Reihe von Sprachwissenschaftlern, die sich mit dem Thema der Repräsentativität beschäftigen.⁷⁹ So schreibt er unter anderem, dass zunächst noch gar nicht von der Repräsentativität von Korpora gesprochen wird, sondern dass Ziff im Jahr 1960 angibt, die aus Korpora ableitbaren Aussagen über die Grundgesamtheit aller in einer Sprache möglichen Äußerungen seien im besten Fall wahrscheinliche Hypothesen, welche jederzeit revidiert werden könnten.⁸⁰ Des Weiteren wird Heger angeführt, der die Kriterien der Belegbarkeit, der Häufigkeit und der Akzeptabilität als unzureichend erklärt, um von den empirisch beobachtbaren Sprachdaten (der *parole*) über die Gesamtheit aller darin aktualisierten Typen (der sogenannten Σ -*parole*) auf das abstrakte System der Strukturregeln (die *langue*) schließen zu können. Für eine Häufigkeitsanalyse, die von der *parole* auf die *langue* schließen möchte, sei eine Extrapolation aus einer exhaustiv analysierten Vorkommensmenge auf die unbekannte Obermenge Σ -*parole* notwendig. Derartige Extrapolationen seien allerdings lediglich im phonetisch-phonologischen Bereich gerechtfertigt, jedoch nicht im lexikalischen aufgrund der Abhängigkeit der Häufigkeiten von thematischen und stilistischen Faktoren.⁸¹ Laut Rieger liege das Problem vor allem darin, dass auf den höheren semiotischen Ebenen derartige Daten über Vorkommensmengen fehlten.⁸² Seiner Meinung nach müssten zunächst zwei Fragen geklärt werden, nämlich zum einen, welche Untersuchungsziele über eine mengenorientiert-textstatistische Analyse an einem vorliegenden

⁷⁹ Vgl. Rieger, Burghard (1979): *Repräsentativität: von der Unangemessenheit eines Begriffs zur Kennzeichnung eines Problems linguistischer Korpusbildung*. Online Publikation: <https://www.uni-trier.de/fileadmin/fb2/LDV/Rieger/Publikationen/Aufsaeetze/79/rub79.pdf> (19.05.2017).

⁸⁰ Vgl. Ziff, Paul (1960): *Semantic Analysis*. Ithaka, NY: Cornell University Press, S. 20.

⁸¹ Heger, Klaus (1970): "Belegbarkeit, Akzeptabilität und Häufigkeit. Zur Aufgabenstellung der Sprachwissenschaft". In: Pilch, H./Richter, H. (Hrsg.) (1970): *Theorie und Empirie in der Sprachforschung*. Basel/ München/ New York: Karger, S. 25.

⁸² Vgl. Rieger (1979): op. cit., S. 8.

Untersuchungsgegenstand sinnvoll angegangen (und gelöst) werden könnten, und zum anderen, welcher Untersuchungsgegenstand zugänglich sein müsse, damit eben diese Analyse sinnvoll angegangen (und gelöst) werden könne.⁸³ In Bezug auf die Repräsentativität der Auswahl sprachlich realisierter Äußerungen, die ein Korpus bilden, schreibt Schaefer, dass eben diese Auswahl dann repräsentativ sein müsse, wenn Klassifizierungen und Verallgemeinerungen über die Sprache und nicht nur bloße Aufzählungen von Korpusbelegen oder deren Nutzung zur Veranschaulichung angestrebt würden. Seiner Meinung nach könnte man dann von Repräsentativität sprechen, wenn der Häufigkeit des Vorkommens aller individuellen Sprachrealisationen entsprechend die jeweiligen Teilmengen aufgrund bestimmter statistischer Kriterien durch die ausgewählten Texte adäquat vertreten wären.⁸⁴ Schank/Schoental zufolge ist die Entscheidung darüber, welche Faktoren des Kommunikationsaktes vernachlässigt, konstant gehalten oder variiert werden können, die Aufgabe des Sprachwissenschaftlers, um Repräsentativität zu erreichen.⁸⁵ Nach dem Zusammentragen all dieser Einschätzungen ist Rieger selbst, wie dem Titel seines Aufsatzes bereits zu entnehmen ist, der Meinung, dass der Begriff der Repräsentativität bestenfalls der Name für das zentrale Problem der Korpusbildung sei, jedoch gewiss nicht dessen Lösung. Seiner Meinung nach gibt es mindestens zwei Schwierigkeiten im Hinblick auf die Repräsentativität eines Korpus. Zum einen soll es angeblich repräsentativ sein, wenn die Stichprobe, welche aus der Grundgesamtheit entnommen wurde, Letztere in Bezug auf bestimmte Parameter je nach deren beobachtetem Vorkommen, anteilmäßig adäquat darstellt, wobei unklar ist, wann die Darstellung noch adäquat ist und wann nicht. Zum anderen ist der Begriff auch deshalb unscharf, weil einige Größen der Bestimmung der Repräsentativität nicht oder nur unzureichend bekannt sind, so beispielsweise die relative Häufigkeit, mit der ein sprachliches Phänomen in einer Stichprobe erscheint, oder die Wahrscheinlichkeit, mit der ein solches Phänomen generell auftritt.⁸⁶ Es gilt hier, dass sich bei zunehmender Korpusgröße die relative Häufigkeit des Merkmals in der Stichprobe immer weiter der Wahrscheinlichkeit annähert, mit der es generell auftritt (die

⁸³ Vgl. Rieger (1979): op. cit., S. 14.

⁸⁴ Vgl. Schaefer, Burkhard (1976): „Maschinenlesbare Textkorpora des Deutschen und des Englischen“. In: *Deutsche Sprache*, 4, S. 359-361.

⁸⁵ Vgl. Schank, Gerd/ Schoenthal, Gisela (1976): *Gesprochene Sprache. Eine Einführung in Forschungsansätze und Analysemethoden*. Tübingen: De Gruyter, S. 17-18.

⁸⁶ Vgl. Rieger (1979): op. cit., S. 11-13.

sogenannte Grenzwertdefinition der Wahrscheinlichkeit), wobei Rieger diese „praktisch nicht einholbare Anweisung“ kritisch sieht, gerade im Hinblick auf die Repräsentativität von Korpora. Daraus ergibt sich nämlich die „paradoxe Folgerung“, dass eine Stichprobe nur dann als repräsentativ gilt hinsichtlich eines Merkmals, wenn über die Grundgesamtheit, aus der sie entnommen wurde, so viel bekannt ist, dass die Erhebung der Stichprobe nicht mehr nötig ist. Da das Wissen über die Grundgesamtheit, also die Wahrscheinlichkeit eines Merkmals, jedoch nur über Stichproben erhoben werden kann, ist die Argumentation zwangsläufig zirkulär. Zudem verdeutlicht die Grenzwertdefinition der Wahrscheinlichkeit, die das ideale, also unter gleichen Bedingungen beliebig oft wiederholbare Zufallsexperiment voraussetzt, dass „die Möglichkeit von Wahrscheinlichkeitsaussagen über und die Notwendigkeit der Zufälligkeitsäußerung an tatsächliche Ereignisfolgen einen zwar in seinen Resultaten erfahrungsgemäß zutreffenden, selbst aber nicht *zirkelfrei definierbaren* Zusammenhang darstellt“. ⁸⁷ Laut Rieger ist das größte Problem bei der Behandlung des Themas repräsentativer Stichproben beziehungsweise Korpora die Tatsache, dass das Kriterium der Zufälligkeit nicht erkannt und für die Korpusbildung nicht gefordert wird. ⁸⁸

Die Repräsentativität als eine der wichtigsten Eigenschaften guter Korpora stellt dabei keine absolute Größe dar, sondern eine relative. Ein Korpus sollte immer repräsentativ sein für den zugrunde liegenden sprachlichen Ausschnitt, wobei es sich um die gesamte Breite der Sprache handeln kann oder um bestimmte diatopisch, diaphasisch, diastratisch oder Genre bezogene Teilbereiche. Scherer führt hierfür das Beispiel der Jugendsprache an. Ein Korpus der Jugendsprache sollte so zusammengestellt sein, dass es allgemeingültige Aussagen über die Jugendsprache ermöglicht. Hierbei müsse zunächst geklärt werden, welche Personen im Hinblick auf ihr Alter als Jugendliche gelten. Außerdem müsse festgelegt werden, unter welchen Bedingungen und in der Kommunikation mit welchen Personen die schriftlichen oder mündlichen Texte von Jugendlichen als Jugendsprache gelten. So seien beispielsweise Schulaufsätze und Unterhaltungen mit Großeltern nicht eindeutig miteinzubeziehen, von Erwachsenen für Jugendmagazine verfasste Artikel aber unter Umständen schon. Intuitive Eingrenzungen

⁸⁷ Rieger (1979): op. cit., S. 13-14.

⁸⁸ Vgl. Rieger (1979): op. cit., S. 15.

müssen also durch konkrete Richtlinien ersetzt werden, welche im Einzelfall bei der Zuordnung helfen.⁸⁹

Auch Teubert/Čermáková sehen diese Problematik bei der Integration verschiedener Textsorten. Sie versuchen den Begriff der Repräsentativität von Korpora von zwei Seiten einzugrenzen. Zunächst wollen sie die Repräsentativität eines Korpus einer bestimmten Sprache über die Eingrenzung der Sprache selbst definieren. Sie stellen sich die Frage, welche Texte für eine bestimmte Sprache repräsentativ sind und daher Teil des Korpus werden sollten. Hierbei halten die Autoren fest, dass eine Sprache aus der Gesamtheit der verbalen Interaktionen besteht, die in der Sprachgemeinschaft in der Vergangenheit stattgefunden haben und in der Zukunft stattfinden werden, also nicht nur veröffentlichte Medien, sondern auch Texte des alltäglichen Lebens wie Briefe, Werbeanzeigen und Arbeitsberichte. Diese Definition bezieht auch Idiolekte, Soziolekte, Dialekte, regionale Varianten, Slang und noch viele andere Formen der Sprache mit ein.⁹⁰

Scherer schreibt diesbezüglich wiederum, es müsse berücksichtigt werden, dass einige Textsorten den Sprechern einer Sprachgemeinschaft häufiger begegnen als andere. So ist der Kontakt mit Zeitungstexten, Fernsehinterviews oder Geschäftsbriefen weitaus frequenter als beispielsweise derjenige mit öffentlichen Vorträgen oder Gerichtsurteilen. Dieses Ungleichgewicht und der Unterschied zwischen den Texten, welche aktiv produziert oder passiv aufgenommen werden, sind ebenfalls bei der Repräsentativität von Korpora, vor allem derer der Standardsprache, zu berücksichtigen. Je nach deren Häufigkeit sind die jeweiligen Textarten nach unterschiedlichen Anteilen in das Korpus aufzunehmen. Dies sei allerdings vom Untersuchungsgegenstand abhängig und es gäbe dafür kein Patentrezept.⁹¹

Somit stellt sich für Teubert/Čermáková die Frage nach der Einteilung in Sprachgemeinschaften, was sich bei genauer Betrachtung ebenfalls als problematisch erweist. Die Sprecher selbst handeln immer wieder neu aus, wer zu ihrer Gemeinschaft gehört und wer nicht. Die Sprachgemeinschaften sind also gesellschaftliche Konstrukte und keineswegs konstante Gruppen. Nachdem die Einteilung in Sprachen und

⁸⁹ Vgl. Scherer (2006): op. cit., S. 5-6.

⁹⁰ Vgl. Teubert, Wolfgang/ Čermáková, Anna (2007): „Directions in corpus linguistics“. In: Teubert, W./ Čermáková, A. (Hrsg.) (2007): *Corpus Linguistics. A Short Introduction*. London: Continuum, S. 59-61.

⁹¹ Vgl. Scherer (2006): op. cit., S. 6.

Sprachgemeinschaften eine scheinbar unlösbare Aufgabe darstellt, versuchen die Autoren das Rätsel der Repräsentativität vom anderen Ende her aufzurollen - dem Korpus. Sie zitieren Sinclairs Definition eines Korpus als „a collection of naturally-occurring language text, chosen to characterize a state or variety of a language“.⁹² Die Texte stellen demnach Muster oder einen Querschnitt der Sprache dar. Hier stellt sich wiederum die Frage, wie diese Muster ausgewählt werden sollen, wobei festgehalten wird, dass ein Korpus die Sprache am besten repräsentiert, wenn die 10.000 häufigsten Lexeme der Sprache auch die 10.000 häufigsten Lexeme des Korpus sind. Es muss aber, egal auf welchem Weg man an die Frage der Repräsentativität eines Korpus herantritt, immer festgelegt werden, was das Korpus repräsentieren soll. Solange man die Sprache, die betrachtet werden soll, nicht bestimmt hat, weiß man weder, welche die 10.000 häufigsten Lexeme sind, noch weiß man, welche Bereiche im Korpus abgedeckt werden sollen. Um aber eine bestimmte Sprache zu einer bestimmten Zeit in einer bestimmten Region genau zu erfassen, muss erst der Zugang zu allen relevanten schriftlichen oder mündlichen Texten gewährleistet sein. Wenn nun aber dieser Zugang gewährt ist, so ist ein repräsentativer Querschnitt in Form eines Korpus laut Teubert/Čermáková nicht mehr von Interesse, da unter diesen Umständen die Gesamtheit der zugängigen Texte für Analysen bevorzugt würde. Da dieser Zustand jedoch als Utopie gilt, macht es für die Autoren keinen Sinn, über Repräsentativität zu diskutieren.⁹³

Außerdem weisen sie darauf hin, dass die Einstellung dazu, was als repräsentativ für eine Sprache in einer Sprachgemeinschaft gilt, über die Jahre wechselt. Auch die Ansichten darüber, welche Texte beispielsweise im Englischunterricht gelehrt werden, ändern sich. Während vor hundert Jahren vor allem die ‚hohe‘ Literatur gelehrt wurde, ist man heute der Überzeugung, dass auch die gesprochene Sprache der Mittelklasse in den Unterricht einfließen sollte. Welche Texte für die Sprachgemeinschaft repräsentativ sind und daher in das Korpus integriert werden sollen, bestimmt die Sprachgemeinschaft. Allerdings betonen die Autoren noch einmal, dass dieses von der Gemeinschaft

⁹² Sinclair, John McHardy (1991): *Corpus, Concordance, Collocation*. Oxford: Oxford University Press, S. 171.

⁹³ Vgl. Teubert/ Čermáková (2007): op. cit., S. 61-65.

akzeptierte Korpus nicht als repräsentatives Korpus bezeichnet werden sollte, sondern als Referenzkorpus.⁹⁴

Auch Biber sieht bereits im Jahr 1994 die Repräsentativität eines Korpus als abhängig von der Bandbreite der Textsorten, wenn er schreibt, „Representativeness refers to the extent to which a sample includes the full range of variability in a population“.⁹⁵ Für ihn ist diese Variation der Textsorten jedoch nur eine Komponente, wobei die Breite der linguistischen Distribution, also der Verteilung linguistischer Eigenschaften, die andere Komponente darstellt.

Immer wieder wird die Frage nach der Größe eines Korpus laut, wenn es um die Repräsentativität geht. Bereits 1992 schreibt Fillmore, der als Sprachwissenschaftler selbst zwischen den beiden Lagern steht – den theoretisch arbeitenden Linguisten, die auch *armchair linguists* genannt werden, einerseits, und den *corpus linguists* andererseits – er könne sich in Bezug auf die Repräsentativität nicht ganz auf die Seite der Korpuslinguisten schlagen (zur genaueren Unterscheidung der beiden Herangehensweisen siehe Kapitel 1.2.1. und 1.2.2.). Wenn Korpuslinguisten in einem Korpus bestimmte wichtige Eigenschaften des Englischen nicht finden, schreiben sie dies der mangelnden Größe des Korpus zu, wobei eben diese Fähigkeit zur Beurteilung, ob ein Korpus zu klein ist, um repräsentativ zu sein, auf der Intuition von Muttersprachlern beruht. Daher gibt es seiner Meinung nach keine Möglichkeit, das intuitive Wissen von Muttersprachlern zu umgehen.⁹⁶

Auch Biber legt dar, dass die meisten Forscher davon ausgehen, die Größe des Korpus sei das signifikanteste Charakteristikum der Repräsentativität, also die Frage danach, wie viele Texte integriert werden und wie viele Lexeme pro Text. Jedoch sei die Stichprobengröße nicht das wichtigste Kriterium, sondern eine genaue Definition der Zielgruppe sowie Entscheidungen bezüglich der Sammlung der Stichproben seien vorrangige Charakteristika.⁹⁷

⁹⁴ Vgl. Teubert/ Čermáková (2007): op. cit., S. 65-66.

⁹⁵ Biber, Douglas (1994): „Representativeness in Corpus Design“. In: Zampolli, A./ Calzolari, N./ Palmer, M. (1994): *Current issues in computational linguistics. In honour of Don Walker*. Pisa: Giardini, S. 378.

⁹⁶ Vgl. Fillmore, Charles J. (1992): „'Corpus linguistics' or 'Computer aided armchair linguistics'“. In: Svartvik, J. (Hrsg.) (1992): *Directions in corpus linguistics. Proceedings of Nobel Symposium 82, Stockholm, 4 - 8 August 1991*. Berlin [u.a.]: Mouton de Gruyter, S. 38.

⁹⁷ Vgl. Biber (1994): op. cit., S. 378.

Selbst im Jahr 2011 ist dies ein strittiger Punkt, wenn Kabatek/Pusch ausführen, dass keine Einigkeit darüber herrscht, ab welcher Größe ein Korpus den Anspruch auf Repräsentativität erheben darf. Sie geben zwar Kriterien wie die Systematizität der Sprachdatensammlung oder die bereits mehrfach erwähnte Register- und Textsortenvarianz an, können die Repräsentativität eines Korpus jedoch nicht definieren.⁹⁸

Es herrscht also Uneinigkeit über die notwendige Größe eines repräsentativen Korpus, doch selbst wenn diese Frage geklärt wäre, stellt sich bereits die nächste Frage, nämlich die nach der Auswahl der Texte für ein Korpus. Diesbezüglich schreibt Biber in seinem viel zitierten Aufsatz zur Vorgehensweise bei der Erstellung von repräsentativen Korpora „Representativeness in Corpus Design“, dass nach der Auswahl einer Zielgruppe eine probabilistische Auswahl der Texte erfolgen sollte. Diese kann auf einer willkürlichen Auswahl beruhen, wobei alle Texte die gleiche Chance haben, ausgewählt zu werden. Eine weitere Möglichkeit ist das *stratified sampling* (geschichtete Stichprobe), bei welchem alle Texte, aus denen gewählt wird, in Subgruppen eingeteilt werden, beispielsweise in Textsorten, damit die spätere Auswahl gerecht auf diese aufgeteilt wird. Innerhalb dieser Subgruppen werden die Texte dann wiederum willkürlich ausgewählt. *Stratified samples* sind immer repräsentativer als *non-stratified samples*, da in ersterem Fall die ausgewählten Schichten in den erwünschten Proportionen repräsentiert werden, und nicht von den willkürlichen Auswahltechniken abhängen.⁹⁹ Generell wird bei willkürlich ausgewählten Stichproben ein größeres Korpus benötigt als beim *stratified sampling*, um genauso repräsentativ zu sein (für eine genaue Auflistung von Bibers Parametern zur Textauswahl siehe Anhang 3).¹⁰⁰

Zur Erstellung der Korpora selbst gibt Biber an, dass die Konstruktion von Korpora in Phasen stattfinden sollte. Die erste Phase sollte aus theoretischen Analysen und Pilotstudien bestehen, erst in der zweiten Phase werden dann Texte gesammelt, mit denen ein Pilotkorpus erstellt wird, welches eine große Bandbreite an Variation aber auch Tiefe in einigen Registern bietet und grammatikalisch getaggt wird. Daraufhin sollten weitere empirische Untersuchungen zur linguistischen Variation anhand des Pilotkorpus durchgeführt werden, um die verschiedenen Designparameter zu modifizieren, gefolgt

⁹⁸ Vgl. Kabatek/ Pusch (2011): op. cit., S. 218-219.

⁹⁹ Vgl. Biber (1994): op. cit., S. 379-380.

¹⁰⁰ Vgl. Biber (1994): op. cit., S. 383.

von einer Revision des Designs. Eine Analyse der Variation ermöglicht es schließlich, diejenigen Register zu ermitteln, bei denen das Korpus noch nicht genügend Variation aufweist. Die zyklische Korpuserstellung stellt so eine phasenweise Vergrößerung und Verbesserung des Korpus dar.¹⁰¹ Leech (2007) kritisiert allerdings, dass dieser Aufsatz von Biber zwar häufig zitiert würde, der Vorschlag zur Korpuserstellung aber seines Wissens nach noch bei keinem Korpus angewandt wurde. Außerdem ist er der Meinung, ein nach diesen Kriterien erstelltes Korpus sei nicht repräsentativ, sondern heterogen.¹⁰²

Für Leech sind neben der Repräsentativität auch die Konzepte der Ausgeglichenheit und der Vergleichbarkeit zu beachten. Seiner Meinung nach ist die Ausgewogenheit ein wichtiges Kriterium für die Repräsentativität eines Korpus. Ein Korpus ist demnach ausgeglichen, wenn die Größe der Subkorpora, welche einzelne Genres oder Register repräsentieren, proportional zur relativen Häufigkeit des Vorkommens dieser Genres im textuellen Universum dieser Sprache als Ganzes ist. Die Ausgeglichenheit eines Korpus ist also abhängig von dessen Proportionalität. Allerdings räumt er ein, dass immer noch unklar ist, wann ein Korpus als ausgeglichen gilt und wann nicht. Generell könne man aber von folgender Daumenregel ausgehen: Je größer ein Korpus ist, und je vielfältiger in Bezug auf Genres und andere Sprachvarietäten, desto ausgeglichener und repräsentativer ist es.¹⁰³

Biber erkannte die Proportionalität als Charakteristikum für die Repräsentativität allerdings nie an, da 90% des Korpus dann aus gesprochenen Texten, die nebenbei nur wenig Variation aufzeigten, bestünde, und beispielsweise Reportagen, akademische Texte und Nachrichtensendungen nur einen äußerst geringen Teil des Korpus einnehmen. Für ihn spielen anstatt dessen die Anzahl der Textproduzenten, die Anzahl der Rezipienten und die Texte selbst eine Rolle.¹⁰⁴

Auch Leech ist der Meinung, die Textproduzenten und -rezipienten sollten bei der Auswahl der Texte für ein Korpus eine Rolle spielen, da es doch einen Unterschied macht, ob ein Text, wie beispielsweise ein Gespräch, nur für einen einzelnen Empfänger bestimmt ist, und andere Texte wie Fernsehsendungen für eine mehr oder weniger große

¹⁰¹ Vgl. Biber (1994): op. cit., S. 377, 400-141.

¹⁰² Vgl. Leech, Geoffrey (2007): „New resources, or just better old ones? The Holy Grail of representativeness“. In: Hundt, M./ Nesselhauf, N./ Biewer, C. (Hrsg.) (2007): *Corpus Linguistics and the Web*. Amsterdam/ New York: Rodopi, S. 134.

¹⁰³ Vgl. Leech (2007): op. cit., S. 136-138.

¹⁰⁴ Vgl. Biber (1994): op. cit., S. 386-387.

Gruppe an Personen. Er schlägt daher vor, bei der Textauswahl einen *Initiator-Text-Receiver Nexus* heranzuziehen, welchen er als *Atomic Communicative Event* (ACE) bezeichnet. Er schreibt, dass eine Radiosendung beispielsweise eine Mio. Menschen hören – hierbei gäbe es einen Text, aber eine Mio. ACEs. Das Problem bei dieser Art der Textauswahl ist aber, dass sie kaum praktikabel ist. Es scheint schon fast unmöglich die Anzahl aller Texte in den verschiedensten textuellen Bereichen zu ermitteln, geschweige denn die Größe der jeweiligen Gruppe an Rezipienten herauszufinden. Dennoch verteidigt Leech diese ACE-Proportionalität als Repräsentativität, da er der Meinung ist, obwohl dieses Konzept nicht zu 100% spezifiziert werden kann und es sich um kein hundertprozentig erreichbares Ziel handelt, sondern um eine Skala, bei der eine Annäherung an das Ideal das Ziel sein sollte, so sollte sich bei der Korpuserstellung doch darum bemüht werden, ein so repräsentatives Korpus wie möglich zu erstellen. Des Weiteren können einige Probleme bei der Ermittlung der Rezipienten durch die Beurteilung erfahrener Korpusersteller ausgeglichen werden.¹⁰⁵

Ein dritter Maßstab für die erfolgreiche Erstellung eines Korpus ist laut Leech die Vergleichbarkeit. Zwei Korpora sind dann vergleichbar, wenn sie sich in genau einem Punkt, beispielsweise der temporalen oder regionalen Herkunft der Texte, stark unterscheiden und ansonsten gleich aufgebaut sind. Die Vergleichbarkeit ist zumindest teilweise von der Repräsentativität abhängig: Nur wenn die Korpora repräsentativ sind, erlauben Vergleichskorpora präzise Vergleiche. Bei der Vergleichbarkeit handelt es sich wie bei der Repräsentativität um eine Skala und nicht um ein Ziel, das vollständig erreicht werden kann. Die Ziele der Vergleichbarkeit und Repräsentativität innerhalb eines Korpus sind allerdings inkompatibel, sie konkurrieren also miteinander. Wenn die Einteilung von Texten möglichst vergleichbar sein soll, so spricht die Gruppierung der Texte gegen die Repräsentativität derselben und umgekehrt. Ein Problem der Vergleichbarkeit ist die Entwicklung der Einteilung in Genres. Durch die unterschiedliche Einteilung der Genres durch die Zeit und in verschiedenen Ländern, wird die Erstellung von Vergleichskorpora mit unterschiedlicher temporaler und lokaler Herkunft erschwert.¹⁰⁶ Er lenkt daher am Ende seines Aufsatzes ein:

„The above discussion of representativeness, balance and comparability might lead the reader to reject these concepts as being ill-defined, problematic, unattainable. My attitude is different from

¹⁰⁵ Vgl. Leech (2007): op. cit., S. 138-140.

¹⁰⁶ Vgl. Leech (2007): op. cit., S. 141-143.

1. Korpora und Korpuslinguistik

this. I have tried to show that these are important considerations, and even if we cannot achieve them 100 per cent, we should not abandon the attempt to define them and achieve them. We should aim at a gradual approximation to these goals, as crucial desiderata of corpus design. It is best to recognize that these goals are not all-or-nothing: there is a scale of representativity, of balancedness, of comparability. We should seek to define realistically attainable positions on these scales, rather than to abandon them altogether.“¹⁰⁷

McEnery/Hardie sehen den Begriff der Repräsentativität ebenso als anzustrebendes Ideal, welches nur selten oder gar nicht erreicht werden kann. In der Realität kann sich diesen Konzepten graduell genähert werden, obgleich nicht klar ist, ob eine hundertprozentige Repräsentativität zu erreichen ist. Sie schreiben, dass Konzepte der Repräsentativität weiterhin kritisch analysiert werden und dies schließlich zu einer besseren Definition derselben führen wird.¹⁰⁸ Ihrer Meinung nach ist vor allem die genaue Bestimmung des zu untersuchenden Sprachausschnitts und die proportional passende Auswahl von Texten oder Sprechern von Bedeutung. Bei einer Untersuchung von Verkaufsbeispielen dürften demnach nicht nur solche in das Korpus aufgenommen werden, die in *Coffee shops* stattfinden, sondern die Ortswahl sollte ausgeglichen sein. Dabei sei auch darauf zu achten, dass die Häufigkeit der Gespräche in den jeweiligen Geschäften die proportionale Häufigkeit der Kaufhandlung in eben denselben widerspiegelt.¹⁰⁹

Wie man an all diesen Ausführungen verschiedener Korpuslinguisten erkennen kann, ist das Konzept der Repräsentativität sehr umstritten. Es ist nicht nur fraglich, wie man ein repräsentatives Korpus erstellen kann, sondern auch, wie die Repräsentativität eines bereits bestehenden Korpus ‚gemessen‘ werden soll. Somit bleibt, um es mit McEnery/Hardie auszudrücken, nur eine möglichst gute Annäherung an die Repräsentativität, mit der Hoffnung, die ausgewählten Texte würden den betreffenden Sprachausschnitt so gut wie möglich widerspiegeln.

Zum Schluss bleibt noch auf Tognini-Bonelli hinzuweisen, die einräumt, dass ein Korpus, idealerweise, zwingend repräsentativ sein sollte, dies in der Praxis jedoch ein *Fait accompli* darstellt und der Linguist in einer späteren Phase, wenn die Ergebnisse der

¹⁰⁷ Leech (2007): op. cit., S. 143-144.

¹⁰⁸ Vgl. McEnery, Tony/ Hardie, Andrew (2012): *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press, S. 10-11.

¹⁰⁹ Vgl. McEnery/ Hardie (2012): op. cit., S. 8.

Datenanalyse bewertet werden, manchmal die Parameter, nach welchen das Korpus erstellt wurde, oder die Typologie der integrierten Texte infrage stellen muss.¹¹⁰

Es wird damit abschließend festgehalten, dass das Ziel der Repräsentativität bei der Erstellung eines Korpus auf jeden Fall angestrebt werden sollte, obgleich es unter anderem mangels einer eindeutigen Definition und eines klaren Konzepts nie vollständig erreicht werden kann. Je nach Art des Korpus spielen vor allem die Abdeckung unterschiedlicher Textarten, die richtige Kombination geschriebener und gesprochener Texte sowie die adäquate Abdeckung von Texten aus den verschiedenen spanischsprachigen Ländern eine wichtige Rolle in Bezug auf spanische Korpora. Ein Korpus, das keine Repräsentativität für den darin dargestellten sprachlichen Ausschnitt fordert, kann keine verlässlichen Daten über eben diesen Ausschnitt hervorbringen.

1.2. Was ist Korpuslinguistik?

Nachdem nun geklärt ist, was ein Korpus ist und wie es am besten erstellt sein sollte, muss noch geklärt werden, worum es sich bei der ‚Korpuslinguistik‘ handelt. Um sich dem Begriff zu nähern, ist es unabdingbar, zunächst auf die unterschiedlichen Herangehensweisen an die Sprachwissenschaft einzugehen.

Michael Hoey brachte seine Meinung ganz unmissverständlich zum Ausdruck, als er sagte, "corpus linguistics is not a branch of linguistics, but the route into linguistics".¹¹¹ Diese Aussage stößt bei vielen Linguisten jedoch auf Missfallen, wie im folgenden Kapitel zu sehen sein wird.

In der Korpus- und Dokumentationslinguistik wird die Einteilung von Sprachwissenschaftlern gern überspitzt dargestellt. So können sprachwissenschaftliche Forscher dieser Ansicht nach auf einer Art Skala verortet werden, an deren Enden zwei extreme Forschertypen angesiedelt sind. Diese versuchen auf unterschiedliche Art und Weise, die Sprache zu beschreiben: Auf der einen Seite gäbe es die Theoretiker und auf der anderen die Empiriker. Der Theoretiker sei vor allem am Sprachsystem, welches allen Sprechern einer Sprache zugrunde liege, interessiert. Dieses allen Menschen gemeinsame Sprachvermögen nenne er *Universalgrammatik*. Er denke, übertrieben formuliert, über

¹¹⁰ Vgl. Tognini-Bonelli (2001): op. cit., S. 58.

¹¹¹ Hoey, Michael: „Remark at TALC 1998“. In: Lüdeling, A. (2004): *VL: Einführung in die Korpuslinguistik*. Online Publikation: https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/lehre/alte_jahrgaenge/ss-2004/vl-einf-korpling/pdf/VLGeschichte.pdf (09.02.2017).

die Sprache nach, sinniere über kuriose Beispiele und lege komplizierte und abwegige Sätze, die aber der Grammatik der jeweiligen Sprache entsprächen, Muttersprachlern vor, um von diesen zu erfahren, ob die jeweiligen Sätze wohlgeformt seien oder nicht. Für ihn zähle also einzig und allein das Urteil kompetenter Sprecher. Der Empiriker hingegen sei am Sprachgebrauch interessiert. Er analysiere reales Sprachmaterial und erforsche damit den tatsächlichen Sprachgebrauch der Sprecher einer Sprache. Er sei vor allem an alltäglichen Sprachphänomenen interessiert und beschäftige sich daher mit enormen Datenmengen, die seine Hypothesen entweder bestätigten oder für ungültig erklärten.¹¹² „Ansonsten hält er sich überwiegend an seinem Computer auf“, so Lemnitzer/Zinsmeister.¹¹³ Dies ist selbstverständlich nicht nur eine Abstraktion, sondern auch eine polemische Darstellung, welche dennoch die Grundpositionen beschreibt, und jeder Wissenschaftler kann sich auf der durch die beiden Extreme eingegrenzten Skala verorten – der eine mehr auf der Seite der Theoretiker, der andere mehr auf der Seite der Empiriker. Abgesehen davon wendet sich weder die Mehrheit der Empiriker von der Theorie ab noch lehnen die Theoretiker jegliche empirischen Belege zur Verifizierung ihrer Theorien ab. Die beiden gegensätzlichen Positionen sind auch keineswegs unvereinbar, wie der englische Sprachwissenschaftler Charles J. Fillmore zeigt. Fillmore, einst ein Theoretiker, oder, wie er selbst es ausdrückt, ein ‚armchair linguist‘, wechselte in den achziger Jahren zu den Empirikern.¹¹⁴

Stefanowitsch gibt in seinem Aufsatz „Quantitative Korpuslinguistik und sprachliche Wirklichkeit“ an, dass einige Sprachwissenschaftler trotz der Verwendung elektronischer Korpora noch der introspektiven Methode folgen, indem sie die vorherige introspektive Rekonstruktion der Daten nun einfach durch die persönliche Auswahl der Daten aus dem Korpus ersetzen. Da diese Auswahl vermeintlich relevanter Daten aber auch Teil der Introspektion ist, kann man hier nicht von Korpuslinguistik im eigentlichen Sinn sprechen. Deshalb muss die Korpuslinguistik, so Stefanowitsch, eine systematische und erschöpfende Untersuchung sprachlicher Phänomene sein, wobei das Adjektiv ‚erschöpfend‘ auf die Berücksichtigung aller im Korpus gefundenen Belege anspielt, und ‚systematisch‘ bedeutet, dass die Analyseschritte nicht beliebig variieren dürfen.¹¹⁵

¹¹² Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 6-7. Und: Scherer (2006): op. cit., S. 8.

¹¹³ Lemnitzer/ Zinsmeister (2010): op. cit., S. 6.

¹¹⁴ Vgl. Fillmore (1992): op. cit., S. 35.

¹¹⁵ Vgl. Stefanowitsch (2005): op. cit., S. 149-150.

Des Weiteren schreibt er, dass Korpuslinguistik heute definitiv mit dem Computer verknüpft sei und erst die Verfügbarkeit von Computern mit ausreichender Rechen- und Speicherkapazität die Erstellung von Häufigkeitslisten, Konkordanzen und Kollokationen ermögliche. Die elektronische Datenverarbeitung ist also aus der Korpuslinguistik nicht mehr wegzudenken, allerdings genüge sie nach Stefanowitschs Meinung nicht als definitorisches Kriterium, da allein die Verwendung elektronischer Korpora einen Sprachwissenschaftler noch nicht zum Korpuslinguisten mache. Er müsse auch korpuslinguistische Techniken anwenden. Stefanowitsch betont jedoch ebenfalls, dass die Korpuslinguistik teilweise von der statistischen oder stochastischen Sprachverarbeitung abgegrenzt werden müsse. Diese Forschungstradition, welche die Korpuslinguistik mit der Computerlinguistik gemein habe, könne zwar bei manchen sprachwissenschaftlichen Fragestellungen herangezogen werden, sie müssten aber sprachwissenschaftlicher Natur sein.¹¹⁶

Tognini-Bonelli charakterisiert den Ansatz der Korpusarbeit als dreilinig: Erstens ist es ein empirischer Ansatz zur Beschreibung des Sprachgebrauchs. Zweitens arbeitet er innerhalb des Rahmens einer kontextuellen und funktionellen Theorie der Bedeutung. Und drittens nutzt er die neuen Technologien. Die Korpuslinguistik kann als empirischer Ansatz gesehen werden, da sie wie alle wissenschaftlichen Untersuchungen an realen Daten durchgeführt wird. Die dabei angewandte Vorgehensweise ist induktiv, da der Linguist von der Analyse realer Daten ausgehend, Hypothesen aufstellt, welche aufgrund des wiederholten Vorkommens der Muster bestätigt werden und zu Generalisationen führen. Bezüglich des Rahmens der kontextuellen und funktionellen Theorie der Bedeutung gibt Tognini-Bonelli an, dass die Bedeutung jeder sprachlichen Einheit, ob es sich nun um ein einzelnes Lexem, eine Phrase oder einen Satz handelt, sowohl vom verbalen als auch vom kulturellen Kontext abhängig ist. In Bezug auf den Input der neuen Technologien schreibt Tognini-Bonelli, dass die moderne Technik als Hilfsmittel den methodologischen Rahmen der Untersuchungen beeinflusst hat, dadurch, dass es diese beschleunigt, systematisiert und in Echtzeit auf immer größere Datenmengen anwendbar macht. Des Weiteren ist die Arbeit mit Korpora nun nicht mehr für einige wenige Spezialisten reserviert, sondern steht allen Sprachinteressierten, ob Lehrer, Schüler, Übersetzer oder Lexikographen offen. Als sich die Korpuslinguistik noch auf die

¹¹⁶ Vgl. Stefanowitsch (2005): op. cit., S. 150-151.

1. Korpora und Korpuslinguistik

Verfügbarkeit einer großen Menge von Texten beschränkte, lag das Hauptaugenmerk auf der zentralen Grammatik, dem Kernvokabular und den allgemeinen Regeln. Nun ermöglicht der Computer eine Einschätzung über kontextuelle Angemessenheit, geographische und soziale Varietät, Idiolekt und Stil.¹¹⁷

Lemnitzer/Zinsmeister sehen das ähnlich. Sie verstehen unter ‚Korpuslinguistik‘ eine Teildisziplin der Linguistik und definieren sie wie folgt:

„Als Korpuslinguistik bezeichnet man die Beschreibung von Äußerungen natürlicher Sprachen, ihrer Elemente und Strukturen, und die darauf aufbauende Theoriebildung auf der Grundlage von Analysen authentischer Texte, die in Korpora zusammengefasst sind. Korpuslinguistik ist eine wissenschaftliche Tätigkeit, d.h. sie muss wissenschaftlichen Prinzipien folgen und wissenschaftlichen Ansprüchen genügen. Korpusbasierte Sprachbeschreibung kann verschiedenen Zwecken dienen, zum Beispiel dem Sprachunterricht, der Sprachdokumentation, der Lexikographie oder der maschinellen Sprachverarbeitung.“¹¹⁸

Für sie ist zunächst also wichtig, dass es sich um natürliche Sprache und damit reale Texte handelt sowie dass auch bei der Korpuslinguistik die Einhaltung bestimmter Prinzipien die Grundvoraussetzung ist. Daher müssen die Ergebnisse von Untersuchungen nachprüfbar oder sogar reproduzierbar sein. Die Ergebnisse müssen also an vergleichbaren Korpora überprüft werden können. Außerdem gewährleistet die gemeinsame Nutzung eines Korpus für verschiedene Untersuchungen, dass Forschungsergebnisse miteinander verglichen werden können. Des Weiteren sollten die Methoden anerkannten wissenschaftlichen Standards entsprechen, welche im Bereich der Korpuslinguistik teilweise noch entwickelt werden müssen. Über die Reichweite und Sicherheit von Aussagen, die auf der Basis von Korpusmaterialien erstellt wurden, muss Klarheit herrschen.¹¹⁹

Der vorliegenden Arbeit soll eben diese Definition von Lemnitzer/Zinsmeister zugrunde gelegt werden, wobei sie um Stefanowitschs und Tognini-Bonellis definitorische Verknüpfung mit dem Computer erweitert werden soll, da die ständige Verfügbarkeit und Zugänglichkeit der Korpora nur so gewährleistet werden kann. Somit wird für die vorliegende Arbeit folgende Definition festgehalten:

Unter Korpuslinguistik versteht man die Beschreibung sprachlicher Äußerungen und die darauf aufbauende Theoriebildung auf der Grundlage von Analysen realer Texte,

¹¹⁷ Vgl. Tognini-Bonelli (2001): op. cit., S. 2-6.

¹¹⁸ Lemnitzer/ Zinsmeister (2010): op. cit., S. 10.

¹¹⁹ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 11.

welche in digitalen Korpora gespeichert und zugänglich sind. Sie muss wissenschaftlichen Prinzipien folgen und wissenschaftlichen Ansprüchen genügen.

1.2.1. Entwicklung der Korpuslinguistik

Die Korpuslinguistik ist eine alte Teildisziplin der Sprachwissenschaft, die bereits im 19. Jahrhundert zu existieren begann, aber erst in den letzten Jahren als sprachwissenschaftliche Methode Beachtung fand. Dies ist zum einen auf den Wechsel der Sprachwissenschaft von Chomskys Generativismus zu empirischen Methoden zurückzuführen und zum anderen auf die wachsende Verfügbarkeit sprachwissenschaftlicher Korpora in elektronischer Form bedingt durch die Errungenschaften der modernen Informatik und Computerwissenschaften.¹²⁰ Somit erlebte die Korpuslinguistik seit den 90er Jahren des letzten Jahrhunderts einen enormen Aufschwung.¹²¹

Hierzu schreibt John Sinclair gleich zu Beginn seines Buchs *Corpus, concordance, collocation* (1991) über die Entwicklung der Korpuslinguistik:

„Thirty years ago when this research started it was considered impossible to process texts of several million words in length. Twenty years ago it was considered marginally possible but lunatic. Ten years ago it was considered quite possible but still lunatic. Today it is very popular.“¹²²

Anhand dieses Zitats ist ersichtlich, wie unglaublich ein solches Unterfangen in den ersten Jahren der Korpuslinguistik gewesen sein musste. Kein Wissenschaftler konnte sich vorstellen, je ein solch großes und vor allem digitales Korpus vor sich zu haben, geschweige denn damit arbeiten zu können.

Die moderne Korpuslinguistik zeichnet sich damit durch die folgenden Kriterien aus: Sie arbeitet empirisch, das heißt, sie analysiert reale Sprachmuster anhand natürlicher Texte. Sie verwendet große Textsammlungen, genannt Korpora, als Grundlage für ihre Analyse. Sie nutzt automatische und interaktive Techniken mithilfe des Computers, und sie basiert auf analytischen und quantitativen Techniken.¹²³

Leech beschreibt diese neue Art der Korpuslinguistik, die computerisierte Korpuslinguistik, wie folgt:

„[...] computer corpus linguistics (henceforth CCL) defines not just a newly emerging methodology

¹²⁰ Vgl. Stefanowitsch (2005): op. cit., S. 147.

¹²¹ Vgl. Koch, Peter/ Oesterreicher, Wulf (2011): *Gesprochene Sprache in der Romania. Französisch, Italienisch, Spanisch*. Berlin/ New York: De Gruyter, S. 37.

¹²² Sinclair, (1991): op. cit., S. 1.

¹²³ Vgl. Alcántara Plá (2007): op. cit., S. 15.

1. Korpora und Korpuslinguistik

for studying language, but a new research enterprise, and in fact a new philosophical approach to the subject. The computer, as a uniquely powerful technological tool, has made this new kind of linguistics possible. So technology here (as for centuries in natural science) has taken a more important role than that of supporting and facilitating research: I see it as the essential means to a new kind of knowledge, and as an 'open sesame' to a new way of thinking about language.”¹²⁴

Er sieht die computerisierte Korpuslinguistik also als ein essentielles Mittel zu einer neuen Art des Wissens und der Forschung sowie als einen Schlüssel zum Erfolg für einen neuen Weg, über Sprache nachzudenken. Der Computer an sich hat für ihn in diesem Kontext mehr Bedeutung als nur die Forschung zu unterstützen, er ermöglicht diesen neuen Zugang erst.

Die Vorreiter auf dem Gebiet der Korpuslinguistik waren die englischsprachigen Sprachwissenschaftler, und man ist sich einig, dass das erste, wenn auch noch nicht maschinenlesbare Korpus der nun vorherrschenden Korpuslinguistik von Randolph Quirk in seinem Projekt *Survey of the English Usage* (SEU) ab 1953 entwickelt wurde. Quirk und sein Team registrierten eine Mio. Lexeme manuell auf Karteikarten und ordneten sie nach grammatischen Kategorien. Im Jahre 1975 wurde es dann von Svartvik auf den Computer übertragen.¹²⁵

Das erste maschinenlesbare Korpus war jedoch das *Brown Corpus* aus dem Jahr 1964, welches von Francis und Kučera von der *Brown University* in Providence, Rhode Island erstellt wurde. Es umfasste eine Mio. Lexeme des Gegenwartsenglischen, die aus verschiedenen Bereichen sorgfältig ausgewählt und organisiert wurden, war einfach zu nutzen und wurde so oft überprüft, bis es fast fehlerfrei war.¹²⁶ Weitere frühe maschinenlesbare Korpora, die den Beginn der Korpuslinguistik markierten, waren das *LOB-Corpus* (Lancaster-Oslo-Bergen) von Johansson et al. aus den 1970er Jahren, Jan Svartviks *London Lund Corpus* aus den späten 1970er Jahren und John Sinclairs *COBUILD Corpus*. Letzteres enthielt 18,3 Mio. Lexeme und wurde für die Auswahl der Einträge des *Collins COBUILD English Language Dictionary* erstellt. Dieses 1987 veröffentlichte Wörterbuch ist somit das erste, welches allein auf Korpusdaten basiert.¹²⁷

¹²⁴ Leech, Geoffrey (1992): „Corpora and theories of linguistic performance“. In: Svartvic, J. (Hrsg.) (1992): *Directions in corpus linguistics. Proceedings of Nobel Symposium 82, Stockholm, 4-8 August, 1991*. Berlin [u.a.]: Mouton de Gruyter, S. 106.

¹²⁵ Vgl. Alcántara Plá (2007): op. cit., S. 15.

¹²⁶ Vgl. Teubert, Wolfgang (2007): „Language and corpus linguistics“. In: Teubert, W./ Čermáková, A. (Hrsg.) (2007): *Corpus Linguistics. A Short Introduction*. London: Continuum, S. 52.

¹²⁷ Vgl. Teubert (2007): op. cit., S. 51-57.

Quirks *Survey of the English Usage* beschäftigte sich damals hauptsächlich mit der Grammatik, doch das wirklich Neue war die Arbeit mit empirischen Daten, was zu diesem Zeitpunkt, als gerade Chomskys *Syntactic Structures* (1957) veröffentlicht worden war und das Augenmerk damit auf die Introspektion legte, äußerst problematisch gesehen wurde und auf wenig Akzeptanz stieß. Während man gehofft hatte, das *Brown Corpus* und das *LOB-Corpus* würden aufschlussreiche Studien zu Grammatik und Vokabular ermöglichen, war bald klar, dass Korpora im Umfang von einer Mio. Lexemen nur einen Bruchteil der Sprache wiedergeben könnten, und man verlor teilweise das Interesse daran. Dennoch wurde das *LOB-Corpus* für Forschungen zur Grammatik und vor allem zur Worthäufigkeit verwendet, vor allem von Geoffrey Leech, Knut Hofland und Stig Johansson. John Sinclair war schließlich der erste, der sein Korpus speziell für lexikalische Forschungen, also zur Erforschung der Bedeutung von lexikalischen Einheiten, beispielsweise für das neu entwickelte Konzept der Kollokation, verwendete (weitere Informationen zum Begriff der ‚Kollokation‘ siehe Kapitel 4.1.1.). Von den 1970ern bis zu den 1990ern verschwanden die Studien mit Zielsetzungen zur Bedeutung von Lexemen gänzlich aus der Korpuslinguistik, da sich in dieser Zeit neue Fragen stellten, beispielsweise nach der Repräsentativität eines Korpus, der Zusammenstellung, der Bedeutung der Größe oder der Rolle von Spezialkorpora. Auch fragte man sich, ob Annotationen die Forschungsergebnisse verfälschten und ob ein allgemeines *tagset* für alle Sprachen erstellt werden sollte (siehe Kapitel 1.4.5.). Zudem wurden zu dieser Zeit die Varietätenforschung und die Frequenzanalyse populär, da mithilfe der Korpora nun alles gezählt werden konnte. Anhand der Forschungen zur Häufigkeit wurden zudem grammatische Regeln und die Sicht auf die Syntax überarbeitet. Auch in den einführenden Werken zur Korpuslinguistik in den 1990er Jahren verzichtete man auf genauere Ausführungen zur Semantik. John Sinclairs *Corpus Concordance Collocation* aus dem Jahr 1991 war die erste Veröffentlichung, die sich ausdrücklich mit der Bedeutung von Lexemen beschäftigte und 2001 kam schließlich Michael Stubbs *Words and Phrases: Corpus Studies of Lexical Semantics* hinzu. Generell basierte die Mehrheit der Forschungen jedoch auf Häufigkeitsanalysen und das Korpus wurde, beispielsweise bei Collins *COBUILD English Language Dictionary* hauptsächlich dafür verwendet, die Entscheidungen der Lexikographen zu rechtfertigen und Beispiele zu finden.¹²⁸

¹²⁸ Vgl. Teubert (2007): op. cit., S. 52-57.

In der spanischen Sprachwissenschaft hat die Korpuslinguistik erst relativ spät Einzug gehalten, es ist jedoch hervorzuheben, dass es im Spanischen eine besondere Tradition international vernetzter Korpusprojekte gibt, welche fast die gesamte Hispanophonie umfassen. Das erste Projekt war das *Proyecto de Estudio Coordinado de la Norma Lingüística Culta de las Principales Ciudades de Iberoamérica y de la Península Ibérica* (PILEI, auch bekannt als *Habla culta*) des in Mexiko arbeitenden Madrider Sprachwissenschaftlers M. Lope Blanch. Ab 1970 wurden Interviews und Unterhaltungen durchgeführt, welche den städtischen Sprachgebrauch der oberen Mittelschicht (und später auch anderer Schichten) in verschiedenen Städten Spaniens und Lateinamerikas festhalten sollten.¹²⁹ Ab den frühen 1970er Jahren erschienen diese dann in Buchform, und seit 1998 gibt es eine Auswahl an Transkriptionen auf CD-ROM.¹³⁰ Ein weiteres nahezu weltumspannendes Vorhaben ist das *Proyecto PRESEEA* (*Proyecto para el estudio sociolingüístico del español de España y de América*), welches 1992/93 entstand. Während das *Proyecto PILEI* ein lexikographisches Ziel verfolgte, hat das *Proyecto PRESEEA* eine soziolinguistische Zielsetzung. Die Informanten werden nach drei Bildungsniveaus und weiteren soziolinguistischen Parametern gruppiert. Außerdem wurden die Transkriptionen dieses Projekts von Anfang an elektronisch gespeichert und erschienen auf CD-ROM.¹³¹ Zudem sollen sie über eine zentrale Projektplattform im Internet verfügbar gemacht werden, wobei es bisher (Stand 2019) nur eine Auswahl an herunterladbaren Korpusdaten auf der Internetseite gibt.¹³² Es handelt sich dabei um ein noch laufendes Projekt, welches in Kapitel 2.8.2.1. genauer beschrieben wird.

Laut Davies/Face existierten bis zum Jahr 2001 keine öffentlich zugänglichen Korpora des Spanischen, die mehr als eine Mio. Lexeme enthielten. In jenem Jahr veröffentlichte die *Real Academia Española* zwei große, repräsentative Korpora mit jeweils mehr als 100 Mio. Lexemen, das *Corpus Diacrónico del Español* (CORDE) und das *Corpus de Referencia del Español Actual* (CREA).¹³³ Ein Jahr später, im Jahr 2002,

¹²⁹ Vgl. Kabatek/ Pusch (2011): op. cit., S. 217.

¹³⁰ Vgl. Samper Padilla, Antonio José (1998): *Macrocorpus de la norma lingüística culta de las principales ciudades del mundo hispánico. Intro.* Las Palmas de Gran Canaria: Univ. de las Palmas. CD-ROM.

¹³¹ Vgl. Kabatek/ Pusch (2011): op. cit., S. 217.

¹³² Vgl. Moreno Fernández, Francisco: *Proyecto para el Estudio Sociolingüístico del Español de España y de América. Presentación.* Online Publikation: <http://preseea.linguas.net/Inicio.aspx> (18.06.2019).

¹³³ Vgl. Davies, Mark/ Face, Timothy L. (2006): „Vocabulary Coverage in Spanish Textbooks: How Representative is it?“. In: Sagarra, N./ Toribio, A. J. (Hrsg.): *Selected Proceedings of the 9th Hispanic Linguistics Symposium*. Somerville, MA: Cascadilla Proceedings Project, S. 133.

erschien die erste Version des *Corpus del Español* von Davies. Im Jahr 2013 kam das *Corpus del Español del Siglo XXI* (CORPES) hinzu und 2016 die neue Version des *Corpus del Español*. All diese Korpora werden im zweiten Kapitel detailliert vorgestellt.

Im Allgemeinen wurden und werden immer noch viele Korpora zu lexikografischen Zwecken entwickelt. Ein Beispiel ist das Korpus *Cumbre* von Sanchez et al. aus dem Jahr 2001, welches zur Erstellung des *Gran Diccionario de Uso del Español Actual* erarbeitet und in Teilen auf CD-ROM veröffentlicht wurde.¹³⁴ Des Weiteren ist für die spanische Sprache eine hohe Zahl an diachronen Korpora zu verzeichnen. In jüngster Zeit werden zudem immer mehr Spezialkorpora veröffentlicht, die aus kleineren Projekten stammen, im Internet jedoch von jedermann genutzt werden können. Es finden sich viele Beispiele hierfür im eben erwähnten zweiten Kapitel zur Kategorisierung der Korpora, vor allem im Bereich der morphosyntaktisch, syntaktisch oder semantisch annotierten Korpora (siehe Kapitel 2.5.) oder der Lernerkorpora (siehe Kapitel 2.2.2.).

1.2.2. Theoretische Linguistik und Korpuslinguistik

Wie bereits erwähnt, ist die Entwicklung und Akzeptanz der Korpuslinguistik eng mit der Unterscheidung zwischen den beiden Strömungen der primär theoretisch arbeitenden und der vordergründig empirisch-beobachtenden Linguistik verbunden. Beide Ansätze, der theoretische wie der empirische, verfolgen dasselbe Ziel: Sie wollen die Sprache beschreiben. Beim theoretischen Ansatz wird allerdings bevorzugt das Sprachsystem untersucht, wohingegen beim empirischen Ansatz der Sprachgebrauch im Zentrum steht.¹³⁵

Die Arbeit mit Korpora wird als empirischer Ansatz gesehen, bei dem reale Daten den Ausgangspunkt darstellen. Die Vorgehensweise ist induktiv, weil von Beobachtungen realer Beispiele auf theoretische Aussagen über die Sprache geschlossen wird.¹³⁶ Unter Induktion wird eine Generalisierung, eine „Ableitung des Allgemeinen aus dem Besonderen“ verstanden.¹³⁷ Grundsätzlich wird jeder Erkenntnistheorie eine wissenschaftliche Methode zugeschrieben. Während der Empirismus mit der Induktion

¹³⁴ Vgl. Sánchez, Aquilino/ Cantos, Pascual/ Simón Granda, José (2001): *Corpus Cumbre lingüístico: selección de 2.000.000 de palabras del lenguaje oral y escrito*. Madrid: SGEL, CD-ROM.

¹³⁵ Vgl. Scherer (2006): op. cit., S. 1.

¹³⁶ Vgl. Tognini-Bonelli (2001): op. cit., S. 2.

¹³⁷ Bußmann, Hadumod (2008): *Lexikon der Sprachwissenschaft*. Stuttgart: Kröner, S. 113-114.

verbunden wird, geht die Deduktion mit dem Rationalismus einher.¹³⁸ Der Begriff Deduktion bezeichnet „eine logische Schlussweise, nach der aus einer oder mehreren akzeptierten Aussagen (= Prämissen) eine weitere akzeptierte Aussage (= Konklusion) abgeleitet werden kann“.¹³⁹ So werden in der Generativen Grammatik, als Ansatz, der dem Rationalismus nahe steht, Hypothesen über das menschliche Sprachwissen formuliert und den sprachlichen Daten kommt die Funktion zu, diese Hypothesen zu falsifizieren. Die Generativisten nutzen hierfür Sprecherurteile, da die Sprecher in Bezug auf ihre Muttersprache ein unbewusstes Sprachwissen haben und aufgrund dieses Wissens intuitiv die Grammatikalität von Sätzen oder morphologischen Strukturen beurteilen können.¹⁴⁰

Während im prechomskyanischen Strukturalismus reale Daten, im Sinne von Saussures *parole* beziehungsweise Coserius *Rede* (siehe auch Kapitel 1.1.1.), als der einzig geeignete Zugang zur linguistischen Untersuchung der Sprache galten und in diesem Rahmen das bereits im vorherigen Unterkapitel erwähnte erste Korpus der nun vorherrschenden Korpuslinguistik, das SEU (*Survey of the English Usage*, 1959), veröffentlicht wurde, führte die Revolution des Generativismus am Ende der fünfziger, vor allem aber in den sechziger und siebziger Jahren dazu, dass die strukturalistischen Methoden und Ziele, inklusive der Arbeit mit Korpora, infrage gestellt wurden. Chomsky führte einen radikalen Wechsel im linguistischen Denken an, der vom strukturellen Empirismus zum generativen Rationalismus, vom Funktionalismus zum Formalismus, von der Beobachtung realer Daten zur Introspektion führte.¹⁴¹

Im Rahmen des generativen Ansatzes steht die Kompetenz, welche universell und angeboren ist, im Mittelpunkt. Das Objekt linguistischer Untersuchungen ist daher das unterliegende Regelsystem, welches die Kreation und Interpretation einer unendlichen Anzahl von Sätzen erlaubt. Aus dieser Perspektive gesehen können Korpora nicht als Quelle linguistischer Forschung akzeptiert werden, da sie kein Beleg für die Kompetenz, sondern für die Performanz sind.¹⁴² Generativisten lehnen Korpora als Quellen

¹³⁸ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 18.

¹³⁹ Bußmann (2008): op. cit., S. 113.

¹⁴⁰ Vgl. Becker, Martin (2013): *Einführung in die spanische Sprachwissenschaft*. Stuttgart: Metzler, S. 37.

¹⁴¹ Vgl. Cruz Piñol, Mar (2012): *Lingüística de corpus y enseñanza del español como 2/L*. Madrid: Arco Libros, S. 29-30.

¹⁴² Vgl. Cruz Piñol (2012): op. cit., S. 30.

linguistischen Wissens ab, da die Sprache produktiv ist. Mit einem begrenzten Wortschatz und einer überschaubaren Zusammenstellung an Regeln ermöglicht es uns unser Sprachvermögen eine unendliche Anzahl an Sätzen hervorzubringen, weshalb ein Korpus als zwangsläufig unvollständig und endlich gesehen wird.¹⁴³

Ab dem Ende der achziger Jahre begann man allerdings die Grenzen der rationalistischen Ansätze und des Generativismus selbst zu sehen, wodurch sich die Einstellung zur Nutzung von Korpora als Informationsquelle für linguistische Studien langsam änderte. Zum einen führten die Fortschritte im Bereich der Informatik dazu, dass immer größere Datenmengen verarbeitet werden konnten und immer weniger Fehler bei der automatischen Bearbeitung gemacht wurden. Zum anderen interessierte man sich zunehmend für die soziale Funktion von Sprache, wodurch situationale Kontexte in den Vordergrund traten. Nur reale Daten, welche der Performanz entstammten, konnten die Sprache in ihrem sozialen Gebrauch widerspiegeln, weshalb das Interesse an Korpora erneut entfachte.¹⁴⁴

Es gibt allerdings auch einige Sprachwissenschaftler, die von der einen Strömung zur anderen wechselten. Ein Beispiel hierfür ist der bereits erwähnte Charles J. Fillmore, der in seinem Aufsatz „'Corpus linguistics' or 'Computer-aided armchair linguistics'“ schreibt, dass diese beiden Ansätze sich nicht unüberwindbar gegenüber stehen und am besten vereint werden sollten.¹⁴⁵ Ein weiterer Sprachwissenschaftler, der von der Seite der Generativen Grammatik zur Seite der Korpuslinguisten wechselte, ist Geoffrey Sampson. Als amerikanischer Student in den späten 1960er Jahren war es für ihn selbstverständlich, dem Lager der Generativisten anzugehören, und der Abgleich grammatikalischer Theorien mit objektiven Sprachbelegen spielte damals keine große Rolle in der Sprachwissenschaft. Während eines Forschungsvorhabens zu *Multiple central embeddings* (wenn beispielsweise in einen Hauptsatz ein Relativsatz integriert ist, welcher seinerseits wiederum eine derartige Erweiterung enthält) entdeckte er, dass die Intuition der Muttersprachler in diesem Fall falsch war und die tatsächliche Verwendung nur anhand von Belegen zum tatsächlichen Sprachgebrauch ermittelt werden konnte. Nach diesem Projekt und einem Sabbatjahr in der Schweiz wechselte Sampson folglich

¹⁴³ Vgl. Teubert (2007): op. cit., S. 46.

¹⁴⁴ Vgl. Cruz Piñol (2012): op. cit., S. 30-31.

¹⁴⁵ Vgl. Fillmore (1992): op. cit., S. 35.

zur Korpuslinguistik, wo er mit Geoffrey Leech zusammenarbeitete, und sah diesen Wechsel von da an als die beste Entscheidung in seiner Karriere.¹⁴⁶

Zudem kann derzeit nicht mehr von einer derart steifen Einteilung in den funktionalen und den formalen Ansatz, den Empirismus und den Rationalismus gesprochen werden, wie noch in den sechziger Jahren. Zum einen stellte sich heraus, dass Korpora ohne die Intuition nicht richtig analysiert und interpretiert werden können. So sind linguistische Untersuchungen, die von realen Sprachproduktionen, also der Performanz, ausgehen, jedoch das Wissen sowie angeborene Prinzipien (die Kompetenz), nicht verleugnen, am erfolversprechendsten. Zum anderen wurden Korpora bereits seit den achziger Jahren auch in generativen Studien zur Aneignung der Muttersprache sowie von Fremdsprachen genutzt, was belegt, dass sogar für die Generative Grammatik Korpora eine wichtige Quelle für die linguistische Forschung darstellen.¹⁴⁷

Alcántara Plá ist der Meinung, das am schwersten zu lösende Problem der Korpuslinguistik läge im unendlichen Charakter der Sprache, und schreibt Folgendes über das Verhältnis zwischen Korpora und Sprache: “Si aceptamos que una lengua permite un conjunto infinito de combinaciones, es claro que no podemos optar a contener en un corpus todos sus casos posibles.”¹⁴⁸ Ein Sprachwissenschaftler, Lehrer und auch ein Schüler muss sich bei der Arbeit mit Korpora immer darüber im Klaren sein, dass ein Korpus nur Beispiele bereitstellen kann, Suchen anhand einer großen Zahl von Beispielen erlaubt und Frequenzanalysen ermöglicht, dass diese Sammlung von Belegen aber nicht vollständig ist und einige tatsächlich existierende Strukturen eventuell nicht enthält. Wie Paul M. Lloyd es ausdrückt: “[...] language is not an object in any physical sense, but rather an activity that is expressed by physical means (i.e., the sounds of language). Each speaker in essence creates language each time he or she speaks [...]”¹⁴⁹ Bei dieser Aussage hat Lloyd sich an Coseriu orientiert, welcher bereits im Jahr 1973 in seinem bis heute zitierten Werk *Sincronía, diacronía e historia* schrieb:

“La lengua cambia justamente porque *no está hecha* sino que *se hace* continuamente por la actividad lingüística. [...] El hablar es actividad creadora, libre y finalista, y es siempre nuevo, en cuanto se

¹⁴⁶ Vgl. Sampson, Geoffrey (1996): „From central embedding to corpus linguistics“. In: Thomas, J./ Short, M. (Hrsg.) (1996): *Using Corpora for Language Research*. London: Longman, S. 14-26.

¹⁴⁷ Vgl. Cruz Piñol (2012): op. cit., S. 32-33.

¹⁴⁸ Alcántara Plá (2007): op. cit., S. 14.

¹⁴⁹ Lloyd, Paul M. (1987): *From Latin to Spanish. Vol. I: Historical Phonology and Morphology of the Spanish Language*. Philadelphia: American Philosophical Society, S. 67.

determina por una finalidad expresiva individual, actual e inédita.”¹⁵⁰

Die Sprache in ihrer ‚Gesamtheit‘ ist also nicht erfassbar, da jeder Sprecher sie mit jeder Äußerung immer wieder neu erschafft. Dieser Unmöglichkeit der Erfassung einer Sprache in einem noch so großen Korpus muss sich der Korpuslinguist und auch jede andere Person, die mit Korpora arbeitet, immer bewusst sein. Die Korpora dienen lediglich als Referenz, als Datenbasis für reale Sprache, die für die Untersuchung einer Einzelsprache genutzt werden, aber nie alle Möglichkeiten sprachlicher Äußerungen umfassen kann.

Lemnitzer/Zinsmeister stimmen dieser Kritik teilweise zu und geben in Bezug auf die Skepsis der Generativisten im Hinblick auf reale Sprachdaten das folgende Beispiel: Wenn man davon ausgeht, dass die Kompetenz anhand der tatsächlich produzierten Äußerungen ersichtlich ist, dann gehören häufig vorkommende, aber grammatikalisch falsche Sätze zum Inventar der Sprache, wohingegen seltene aber richtige Äußerungen nicht dazugehören. Man muss bei einer solchen Vorgehensweise also äußerst vorsichtig sein und Entscheidungen bezüglich der Häufigkeit von denen der Grammatikalität unterscheiden. Daher gelten in der Generativen Grammatik nur Sprecherurteile, also Selbstauskünfte kompetenter Sprecher über gewisse Äußerungen, als akzeptierte Quelle für Rückschlüsse auf die Kompetenz.¹⁵¹ Dennoch können anhand der Korpusdaten zusätzliche Einblicke in den tatsächlichen Sprachgebrauch der Sprechergemeinschaft gewonnen werden, und diese tatsächliche Nutzung der Sprache ist vor allem im Hinblick auf den Fremdsprachenerwerb nicht zu unterschätzen.

Alcántara Plá schreibt hierzu: „El corpus nos ayudará sólo como referencia relativa, pero será fundamental para saber qué fenómenos son frecuentes y cuales marginales, y para así focalizar nuestros esfuerzos en los problemas adecuados.”¹⁵² Genau diese Möglichkeit, die Frequenz der Lemmata und damit die Relevanz des jeweiligen Lexems in der Kommunikation der Muttersprachler zu ermitteln, sollte jedoch auf jeden Fall genutzt werden, sofern man sich der Tatsache bewusst ist, dass die Korpora nur Beispiele und keineswegs die ‚gesamte‘ Sprache, welche als solches ohnehin in keiner Art und Weise festgehalten werden kann, enthalten.

¹⁵⁰ Coseriu, Eugenio (1973): *Sincronía, diacronía e historia. El problema del cambio lingüístico*. Madrid: Gredos, S. 69.

¹⁵¹ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 22.

¹⁵² Alcántara Plá (2007): op. cit., S. 14.

1. Korpora und Korpuslinguistik

Darüber hinaus stellt sich für Tognini-Bonelli die Zuteilung der Korpuslinguistik zur angewandten Linguistik nicht ganz so klar dar. Sie schreibt, dass sich die Korpuslinguistik, obgleich sie zur angewandten Linguistik zählt, doch sehr von den anderen Disziplinen in dieser Kategorie unterscheidet, da sie als *pre-application methodology* gesehen werden kann. Unter einer Methodologie versteht sie die Anwendung einer gegebenen Anzahl von Regeln oder Wissensbeständen in einer bestimmten Situation. Im Gegensatz zu anderen Herangehensweisen, welche damit beginnen, gewisse Tatsachen als gegeben zu betrachten, definiert die Korpuslinguistik ihre eigenen Regeln und Wissensbestände, bevor diese angewandt werden. Folglich verwendet der Linguist einige neue Parameter, um die Daten zu erklären, was wiederum zu einer Veränderung in der ‚Einheit der Währung‘ für die sprachwissenschaftliche Beschreibung führt. Die Korpuslinguistik hat daher einen theoretischen Status und ist somit in der Lage, zum Fortschritt auf anderen Gebieten beizutragen.¹⁵³

1.2.3. Korpuslinguistik: Methode oder Disziplin?

Eine Frage, die sich im Bereich der Sprachwissenschaft immer wieder stellt, ist diejenige, ob die Korpuslinguistik nun als Methode oder als Disziplin zu betrachten ist. Über die Entwicklung der Korpuslinguistik hinweg haben sich Sprachwissenschaftler und insbesondere Korpuslinguisten immer wieder diese Frage gestellt und sind bisher zu keiner übereinstimmenden Meinung gekommen. Vielmehr teilt die Antwort auf eben diese Frage die Korpuslinguisten in zwei Lager.

Laut Burr (2004) war die Korpuslinguistik zunächst keine neue Disziplin, sondern eine spezifische Methodologie, die in verschiedenen Bereichen der Linguistik zum Tragen kommt.¹⁵⁴ McEnery/Wilson definieren sie ebenso als eine Technik, Herangehensweise oder Methodologie: „Corpus linguistics is a methodology that may be used in almost any area of linguistics, but it does not truly delimit an area of linguistics itself.“¹⁵⁵ Auch für Anatol Stefanowitsch¹⁵⁶ und Carmen Scherer¹⁵⁷ kann und will die Korpuslinguistik knapp zehn Jahre später nur eine von mehreren empirischen Methoden

¹⁵³ Vgl. Tognini-Bonelli (2001): op. cit., S. 1.

¹⁵⁴ Vgl. Burr, Elisabeth (2004): „Das Korpus romanischer Zeitungssprachen in Forschung und Lehre“. In: Dahmen, W./ Holtus, G./ Kramer, J./ Metzeltin, M./ Schweickard, W./ Winkelmann, O. (Hrsg.) (2004): *Romanistik und neue Medien*. Romanistisches Kolloquium XVI (=TBL 455). Tübingen: Narr, S. 135.

¹⁵⁵ McEnery/ Wilson (1996): op. cit., S. 2.

¹⁵⁶ Vgl. Stefanowitsch (2005): op. cit., S. 159.

¹⁵⁷ Vgl. Scherer (2006): op. cit., S. 2.

innerhalb der Sprachwissenschaft sein.

Tognini-Bonelli hingegen ist der Meinung, dass die Korpuslinguistik, trotz der anfänglichen Feststellung, sie sei lediglich eine methodologische Basis für das Sprachstudium, weit über diese Rolle hinausgeht. Sie sieht in der derzeitigen Entwicklung der Korpuslinguistik ein neues Forschungsunternehmen und einen neuen philosophischen Ansatz in Bezug auf die linguistische Forschung.¹⁵⁸ Für sie ist die Korpuslinguistik eine Disziplin, welche die methodologische Basis einiger Anwendungen beeinträchtigt und zugleich das linguistische Objekt, das angewandt wird, verändert hat.¹⁵⁹ Auch Kabatek/Pusch bezeichnen die Korpuslinguistik als eine Disziplin, die sich mit digitalen Sammlungen sprachlicher Daten beschäftigt und daher eng mit der Computertechnologie sowie der Möglichkeit der Speicherung und Verarbeitung großer Datenmengen verbunden ist. Ihrer Meinung nach handelt es sich um eine neue Disziplin, da sie in ihrer heutigen Form nicht ohne den technischen Fortschritt möglich gewesen wäre. Diese habe sich allerdings aus einer traditionellen sprachwissenschaftlichen Methode heraus entwickelt – der Beschreibung linguistischer Phänomene anhand deren Vorkommen in Textsammlungen.¹⁶⁰ Auch Lemnitzer/Zinsmeister bestätigen diese Entwicklung und schreiben bereits im Klappentext zu ihrer Einführung in die Korpuslinguistik: „Die linguistische Arbeit mit digitalen Textsammlungen hat sich in den letzten Jahren von einer Methode hin zu einer eigenen Disziplin der Linguistik entwickelt.“¹⁶¹

McEnery/Hardie kritisieren diese Sichtweise von Tognini-Bonelli und einigen anderen (unter ihnen Stubbs, Teubert und Sinclair) allerdings, welche davon ausgehen, dass die Korpuslinguistik einen theoretischen Status habe und der ausschließliche Gebrauch von Korpora zur Entwicklung neuer Sprachtheorien führe. Ihrer Meinung nach würden die Daten und die Theorie dadurch eins, und es würde sogar so weit gegangen, zu sagen, dass das Korpus selbst die Theorie sei.¹⁶² Ein Beispiel für diese Sichtweise ist Tognini-Bonellis Aussage „The theory has no independent existence from the evidence“.¹⁶³ Nach Ansicht dieses Lagers geht also alle Erkenntnis vom Korpus aus, es darf keine zuvor formulierte Theorie angewandt werden. McEnery/Hardie bezeichnen

¹⁵⁸ Vgl. Tognini-Bonelli (2001): op. cit., S. 1.

¹⁵⁹ Vgl. Tognini-Bonelli (2001): op. cit., S. 9.

¹⁶⁰ Vgl. Kabatek/ Pusch (2011): op. cit., S. 208.

¹⁶¹ Lemnitzer/ Zinsmeister (2010): op. cit., Klappentext.

¹⁶² Vgl. McEnery/ Hardie (2012): op. cit., S. 147.

¹⁶³ Tognini-Bonelli (2001): op. cit., S. 84.

diesen Ansatz daher als *corpus-as-theory*. Sie selbst vertreten jedoch die Tradition *corpus-as-method*. Während das Lager *corpus-as-theory* jegliche Erklärung ablehnt, die nicht direkt aus der Interaktion mit den Daten abgeleitet wird, geht der Ansatz *corpus-as-method* davon aus, dass Korpora Quellen empirischer Daten sind, welche dazu verwendet werden können, um Erklärungstheorien zu unterstützen oder abzulehnen, selbst wenn die Theorie an sich, als Ganzes oder in Teilen, nicht aus den Korpusdaten abgeleitet wurde. Laut McEnery/Hardie ist der Übergang hier jedoch fließend, denn ein Forscher der Richtung *corpus-as-method* kann seine zuvor aufgestellte Theorie ebenso nach der Analyse der Korpusdaten anpassen, also vielmehr im Sinne von *corpus-as-theory*.¹⁶⁴ Burr zufolge kommt heute in der Korpuslinguistik auch die Intuition der Sprechenden einer Sprache und deren Fähigkeit, die Daten zu interpretieren, zum Tragen sowie das Wissen, das Linguisten über die Sprache haben. Gleichzeitig gibt Burr an, dass sich die Korpuslinguistik seit Beginn der 90er Jahre als eigenständiger Zweig der Linguistik versteht, obgleich sie auch weiterhin zum Teil nicht als Disziplin, sondern als Methode gesehen wird.¹⁶⁵ Klöden ist der Meinung, dass die Korpuslinguistik gerade erst dabei ist, sich als selbstständige Disziplin zu etablieren.¹⁶⁶

Diese Diskussion könnte ewig fortgeführt werden und sie kann und soll in dieser Arbeit nicht abschließend geklärt werden. Dennoch lässt sich über die Jahre hinweg eine gewisse Tendenz zur Bezeichnung der Korpuslinguistik als Disziplin erkennen, was ihren Stellenwert erhöht und im Allgemeinen die Nutzung der Korpora vorantreibt. Obgleich man nicht, wie beispielsweise Tognini-Bonelli¹⁶⁷, davon ausgehen muss, dass jegliche Erkenntnis zwangsläufig dem Korpus entspringen muss und Theorien, die vor der Arbeit mit Korpora aufgestellt wurden, ungültig sind, ist die Korpuslinguistik heute doch mehr als eine bloße Methode.

Generell richtet sich der Stellenwert der Korpuslinguistik für die eigene Forschung auch nach der Herangehensweise und Nutzung der Korpora. Wie bereits erwähnt wurde und im folgenden Unterkapitel noch genauer zu sehen sein wird, ist die Bedeutung der

¹⁶⁴ Vgl. McEnery/ Hardie (2012): op. cit., S. 148-150.

¹⁶⁵ Vgl. Burr (2004): op. cit., S. 138 & 135.

¹⁶⁶ Vgl. Klöden, Hildegard (2002): „Romanistische Korpuslinguistik: Leistungen, Grenzen, Perspektiven“. In: Pusch, C./ Raible, W. (Hrsg.) (2002): *Romanistische Korpuslinguistik. Romance Corpus Linguistics*. Korpora und gesprochene Sprache. Corpora and Spoken Language (= ScriptOralia 126). Tübingen: Narr, S. 7-8.

¹⁶⁷ Vgl. Tognini-Bonelli (2001): op. cit., S. 84.

Korpuslinguistik und damit ihre Einordnung als Methode oder Disziplin stark von der wissenschaftstheoretischen Ausrichtung des jeweiligen Forschers und seiner Zugehörigkeit zu einer der linguistischen Denkschulen sowie deren praktiziertem Ansatz bei der Arbeit mit Korpora abhängig.

1.3. Ansätze der Korpuslinguistik

Den eben beschriebenen Richtungen wird jeweils ein Ansatz zugeschrieben. Die Vertreter von *corpus-as-method* nutzen den *corpus-based approach*, während bei *corpus-as-theory* der *corpus-driven approach* angewandt wird.¹⁶⁸ Auch in vielen deutschen Werken werden englische Begriffe zur Bezeichnung derjenigen Ansätze verwendet, denen bisher kein deutscher Begriff entspricht.¹⁶⁹ In der korpuslinguistischen Forschung der einzelnen Sprachen herrscht jedoch keine Einigkeit über die Bezeichnung der einzelnen Ansätze, weshalb es vor allem bei den vorherrschenden Begriffen der deutschen und der englischen Forschung zu Missverständnissen kommen kann. Das, was unter dem *corpus-based approach* verstanden wird, wird in der deutschen Forschung, beispielsweise bei Lemnitzer/Zinsmeister, als korpusgestützter Ansatz bezeichnet, während das Konzept des *corpus-driven approach* als korpusbasierter Ansatz betitelt wird.¹⁷⁰ Hinzu kommt, dass ursprünglich englische *corpus-based studies* in der deutschen Übersetzung meist als korpusbasierte Studien beschrieben werden, während *corpus-driven studies* als korpusgeleitet oder korpusgestützt übersetzt werden. Die Mehrzahl der Studien wird jedoch als *corpus-based* beziehungsweise in der Übersetzung als korpusbasiert bezeichnet, weshalb dieser Begriff der geläufigste ist.

1.3.1. *Corpus-based approach* oder korpusgestützter Ansatz

Bei der Herangehensweise, die sich als *corpus-based* bezeichnet, wird das Sprachmaterial der Korpora als Quelle zur Überprüfung zuvor formulierter Theorien verwendet.¹⁷¹ Die Beweise aus dem Korpus werden hauptsächlich als Beispiele genutzt, um vorhandene theoretische Aussagen zu erläutern, zu testen oder zu exemplifizieren.¹⁷²

¹⁶⁸ Vgl. McEnery/ Hardie (2012): op. cit., S. 150.

¹⁶⁹ Vgl. Gerstenberg, Annette (2013): *Arbeitstechniken für Romanisten. Eine Anleitung für den Bereich Linguistik*. 2. Auflage. Berlin/ Boston: De Gruyter, S. 86.

¹⁷⁰ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 32-38.

¹⁷¹ Vgl. Gerstenberg (2013): op. cit., S. 86-87.

¹⁷² Vgl. Tognini-Bonelli (2001): op. cit., S. 10-11.

Die Korpusdaten selbst widerlegen diese Aussagen nicht, obwohl sie vielleicht ein ganz anderes Bild darstellen.¹⁷³

Dieser Ansatz geht mit dem Generativismus einher und bezeichnet die Verwendung eines Korpus zur Überprüfung und Korrektur theoretischer Aussagen im Rahmen des Rationalismus.¹⁷⁴ Während viele Vertreter des Generativismus die Verwendung von Korpora generell ablehnen, da diese lediglich den Gebrauch der Sprache, also die Performanz, widerspiegeln, man dadurch jedoch nicht auf die Kompetenz schließen kann, gibt es heute doch einige generative Linguisten, die mit Korpora arbeiten, um Hypothesen zu überprüfen oder auch zu falsifizieren. Das Korpus dient hier als Arbeitsinstrument im Sinne der empirischen Auseinandersetzung mit linguistischen Theorien und verbindet damit die linguistische Theoriebildung, also die theoretische Linguistik, und die Korpuslinguistik als eine Methode der Weiterentwicklung linguistischer Theorien (zur Falsifizierung, Modifizierung oder Exemplifizierung).¹⁷⁵ Tognini-Bonelli schreibt hierzu, dass in diesem Kontext das Potential der Korpusbelege nicht vollständig ausgeschöpft wird, da der Reichtum des Sprachgebrauchs, um einige bestehende theoretische Positionen nicht dramatisch zu bedrohen, in vielerlei Hinsicht geopfert wird und ihm nicht gestattet ist, die deskriptiven und theoretischen Aussagen, die diesen idealerweise erklären sollten, zu formen.¹⁷⁶

Wie bereits erwähnt, bezeichnen Lemnitzer/Zinsmeister diesen rein rationalistischen, die Korpusdaten, wenn überhaupt, nur als Beweis für bereits existierende Theorien verwendenden Ansatz als ‚korpusgestützt‘. Weiter führen sie aus, dass aus der Sicht der Generativisten nur Daten aus linguistisch annotierten Korpora oder Belegsammlungen und nicht aus Korpora in Rohform analysiert werden sollten.¹⁷⁷

Dieser von der Generativen Grammatik bevorzugte Ansatz dient meist dazu, nach syntaktischen Konstruktionen zu suchen, die einer bestimmten Theorie folgen, um eben diese Theorie zu bestätigen oder zu widerlegen. Die Tatsache, dass diese meist äußerst komplexen Konstruktionen mit hoher Variabilität die meisten Korpusabfragesprachen überfordern, führt dazu, dass die Benutzung eines Korpus hier, um es mit den Worten von

¹⁷³ Vgl. Tognini-Bonelli (2001): op. cit., S. 17.

¹⁷⁴ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 19.

¹⁷⁵ Vgl. Becker (2013): op. cit., S.198-199.

¹⁷⁶ Vgl. Tognini-Bonelli (2001): op. cit., S. 10-11.

¹⁷⁷ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 34.

Lemnitzer/Zinsmeister auszudrücken „der Suche nach einer Nadel im Heuhaufen“ gleicht.¹⁷⁸ Dies trägt freilich weiter zur mangelnden Akzeptanz der Korpora seitens der Generativen Grammatik bei.¹⁷⁹

Die Begriffe *corpus-based* und ‚korpusgestützt‘ entsprechen sich gegenseitig jedoch auch nicht zu 100 Prozent, da laut Gerstenberg der Begriff ‚korpusgestützt‘ immer dann verwendet werden soll, wenn ein Ansatz weder als *corpus-based* noch als *corpus-driven* bezeichnet werden kann. Sie gibt dabei an, diesen Begriff zu gebrauchen, um Verwechslungen bei weiteren Ansätzen zu vermeiden.¹⁸⁰

1.3.2. Corpus-driven approach oder korpusbasierter Ansatz

Bei der Methode, die als *corpus-driven* bezeichnet wird, werden aus den im Korpus gefundenen Daten neue Fragestellungen und Theorien entwickelt.¹⁸¹ Die Beobachtung bestimmter Muster führt zu einer Hypothese, die wiederum zu einer Generalisierung in Bezug auf Gebrauchsregeln und schließlich zu einer theoretischen Aussage führt.¹⁸² Die Daten des Korpus werden hier also bereits zum Aufstellen von Hypothesen genutzt. Bei diesem Ansatz, der eng mit dem Kontextualismus verbunden ist, sind die Korpusdaten die primäre Quelle der Erkenntnis, und man geht bewusst nicht über das aus den Daten Generalisierbare hinaus. Der Korpuslinguistik kommt hier also nicht nur die Rolle einer Arbeitsmethode zu, sondern sie wird als eigene Theorie eingestuft, denn die Korpusdaten werden als Grundlage von Verallgemeinerungen gesehen, mit dem Ziel einer einheitlichen Theoriebildung.¹⁸³ Es wird behauptet, dass dieser Ansatz eine qualitative Veränderung in der Sprachbeschreibung mit sich bringt und einige bedeutende Annahmen der zugrunde liegenden traditionellen Linguistik erschüttert.¹⁸⁴

Diese empirisch orientierte Forschung, bei der aus Beobachtungen an realem Material Gesetzesaussagen abgeleitet werden, welche durch weitere Beobachtungen bestätigt, modifiziert oder verworfen werden, wird bei Lemnitzer/Zinsmeister ‚korpusbasierte Linguistik‘ genannt.¹⁸⁵ Sie unterscheiden innerhalb der korpusbasierten

¹⁷⁸ Lemnitzer/ Zinsmeister (2010): op. cit., S. 38.

¹⁷⁹ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 37-38.

¹⁸⁰ Vgl. Gerstenberg (2013): op. cit., S. 87.

¹⁸¹ Vgl. Gerstenberg (2013): op. cit., S. 87.

¹⁸² Vgl. Tognini-Bonelli (2001): op. cit., S. 17.

¹⁸³ Vgl. Becker (2013): op. cit., S. 199.

¹⁸⁴ Vgl. Tognini-Bonelli (2001): op. cit., S. 11.

¹⁸⁵ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 19.

Linguistik zwei verschiedene Ansätze: den *korpusbasierten, quantitativen Ansatz* und den *korpusbasierten, quantitativ-qualitativen Ansatz*.

Generell wird bei der qualitativen Korpusanalyse das Augenmerk auf die Ermittlung, Klassifizierung, Einordnung und Interpretation bestimmter linguistischer Phänomene gelegt, wohingegen bei der quantitativen Analyse vor allem die Frequenz der Phänomene ermittelt und miteinander verglichen wird.¹⁸⁶

1.3.2.1. Der korpusbasierte, quantitative Ansatz

Beim korpusbasierten, quantitativen Ansatz werden auf der Basis von rohen, also nicht linguistisch annotierten Korpora quantitative Daten ermittelt, ohne diese zu interpretieren. Diese Herangehensweise ist extrem empirisch und wird vor allem bei der Informationserschließung und der Verarbeitung gesprochener Sprache angewandt. Bei diesem Ansatz gibt es zwei quantitative Verfahren mittels derer Informationen gesucht und zusammengestellt werden können.¹⁸⁷

Zum einen können N-Gramme mit Frequenzen mittels der N-Gramm-Analyse ermittelt werden. Bei N-Grammen handelt es sich um Folgen von ein, zwei, drei oder mehr Phonemen oder Graphemen, nach denen Texte durchsucht werden. Bei der N-Gramm-Analyse werden somit nicht nur alle Vorkommen des gesuchten Lexems als Treffer angezeigt, sondern auch formähnliche Lexeme. Lemnitzer/Zinsmeister führen als Beispiel an, dass bei der Suche nach dem Lexem *Herzoperation* auf diese Weise auch der Ausdruck *Operation am offenen Herzen* gefunden wird. Hier finden sich beispielsweise acht Trigramme, welche beide Ausdrücke gemeinsam haben (*_He, Her, erz, per, era, ati, tio, ion*), also 90 Prozent der Trigramme des Suchausdrucks. Mit der N-Gramm-Analyse können somit auch andere Schreibvarianten der gesuchten Ausdrücke gefunden werden.¹⁸⁸

Zum anderen kann eine latent-semantische Analyse vorgenommen werden, mithilfe derer auch bedeutungsähnliche Lexeme gefunden werden. Das Ergebnis einer solchen Analyse ist eine Term-Dokument-Matrize, in der jedes in den Dokumenten vorkommende Lexem eine Zeile einnimmt und jeder Text eine Spalte. Durch einen Vergleich der in den einzelnen Dokumenten vorkommenden Terme kann ermittelt

¹⁸⁶ Vgl. Scherer (2006): op. cit., S. 36.

¹⁸⁷ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 34.

¹⁸⁸ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 34-35.

werden, welche Lexeme einem bestimmten Thema angehören und welche nicht. So ist es möglich, einen Text über Fahrräder, in welchem das Lexem *Nabe* nicht erscheint, dennoch als Treffer anzuzeigen, wenn exakt nach diesem Lexem gesucht wird. Allein das häufige gemeinsame Vorkommen des Begriffs *Nabe* mit anderen Lexemen aus diesem Bereich, wie *Speiche* oder *Felge*, ermöglicht die korrekte Zuordnung auf der Basis der semantischen Ähnlichkeit.¹⁸⁹

Mittels dieses Ansatzes können also die absolute und relative Häufigkeit von Lexemen, deren Rangplatz und Distribution, sowie die semantische Ähnlichkeit von Lexemen, gemessen an der Häufigkeit ihres Kovorkommens mit anderen Lexemen, ermittelt werden.¹⁹⁰

1.3.2.2. Der korpusbasierte, quantitativ-qualitative Ansatz

Dieser dem Kontextualismus verpflichtete Ansatz ist dem vorherigen sehr ähnlich. Auch hier bildet das Korpus in Rohform die Grundlage aller Untersuchungen. Experimente und Sprecherbefragungen werden als Quellen ausgeschlossen, denn das Korpus ist die ausschließliche Basis für linguistische Untersuchungen. Allerdings ist dieser Ansatz nur gemäßigt empiriebasiert und die Daten werden im Gegensatz zum korpusbasierten, quantitativen Ansatz von den Belegen ausgehend interpretiert. Zur Interpretation werden zum Teil grammatische Kategorien herangezogen, welche nicht aus den Daten selbst abgeleitet wurden. Dieser Ansatz wird vor allem für Untersuchungen über Kollokationen und semantische Prosodien genutzt.¹⁹¹

Diese Art der Korpuslinguistik ist vor allem für die Lexikographie, die Übersetzungswissenschaft und den Sprachunterricht von Nutzen.¹⁹²

Wie es bereits bei der Frage danach, ob die Korpuslinguistik nun eine Methode oder eine Theorie sei, der Fall war, ist es auch bei der binären Unterscheidung zwischen *corpus-based* und *corpus-driven* so, dass der Übergang fließend ist. Nach der Meinung von McEnery/Hardie unterscheiden sich die beiden Lager nämlich nicht in der Nutzung der Korpora zur Entwicklung von Theorien oder als bloße Quelle für Beispiele, da es hier

¹⁸⁹ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 35-37.

¹⁹⁰ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 33.

¹⁹¹ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 37.

¹⁹² Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 37.

viele Zwischenformen gibt, sondern in den unterschiedlichen Standpunkten zum konzeptuellen Status des Korpus beziehungsweise der Korpuslinguistik – also der Frage danach, ob sie einen theoretischen Status hat oder eine linguistische Methodologie ist.¹⁹³

Aufgrund der uneinheitlichen Verwendung der Begriffe *corpus-based*, *corpus-driven*, ‚korpusgestützt‘ und ‚korpusbasiert‘ wird in der vorliegenden Arbeit wie bei McEnery/Hardie verfahren. Sie erklären zwar die eben angeführte Unterscheidung in *corpus-based* und *corpus-driven* nach Tognini-Bonelli, merken jedoch an, dass für sie eine derartige Dichotomie nicht existiert. Ihrer Meinung nach hat ein Korpus keine eigene Theorie der Sprache, wie vom *corpus-driven approach* angenommen, und deshalb lehnen sie auch die grundlegende, binäre Unterscheidung der beiden Ansätze ab. Aus ihrer Sicht ist daher jegliche korpuslinguistische Studie als *corpus-based* zu bezeichnen.¹⁹⁴ Hinzu kommt, dass laut McEnery/Hardie bis zu Tognini-Bonellis starrer Einteilung in *corpus-based* und *corpus-driven* auch Forscher, die sich selbst und ihre Studien nie auf dieser Seite der Dichotomie sehen würden, ihren Ansatz als *corpus-driven* deklarierten, so zum Beispiel Biber¹⁹⁵. Vielmehr wurde zuvor jeglicher induktiver, bottom-up Ansatz, welcher sich auf rohe Korpusdaten stützte, als *corpus-driven* bezeichnet.¹⁹⁶ Um also Unklarheiten sowohl in Bezug auf die generell uneinheitliche Begriffsverwendung als auch im Hinblick auf die deutsche Beschreibung englischsprachiger Studien in dieser Arbeit zu vermeiden, sollen Studien, die anhand eines Korpus durchgeführt wurden, einheitlich als ‚korpusbasiert‘ bezeichnet werden. Dieser Begriff ist zudem der geläufigste, wie in der Einleitung zu diesem Unterkapitel erwähnt, und tendiert daher am wenigsten zu Verwechslungen.

1.4. Verfahren und Werkzeuge der Korpuslinguistik

Um nun mit einem Korpus arbeiten zu können, ist es grundlegend, von den Verfahren und Werkzeugen für die Korpusaufbereitung und -analyse Kenntnis zu haben. Hierbei ist es zunächst unerheblich, welchen der oben erklärten Ansätze der Korpuslinguist verfolgt. Bei einigen Ansätzen werden bestimmte Methoden häufiger

¹⁹³ Vgl. McEnery/ Hardie (2012): op. cit., S. 151.

¹⁹⁴ Vgl. McEnery/ Hardie (2012): op. cit., S.S. 5-6.

¹⁹⁵ Vgl. Biber, Douglas (2009): „A corpus-driven approach to formulaic language in English: multi-word patterns in speech and writing“. In: *International Journal of Corpus Linguistics* 14 (3), S. 275-311.

¹⁹⁶ Vgl. McEnery/ Hardie (2012): op. cit., S. S. 150.

verwendet, bei anderen werden lediglich die Rohdaten berücksichtigt, doch über die folgenden Begriffe und Vorgehensweisen sollte der Nutzer informiert sein.

1.4.1. *Type-token-Relation* und Frequenzlisten

Bei der Arbeit mit maschinenlesbaren Korpora ist die automatische Erkennung von Wortformen besonders bedeutsam, wobei automatisch erstellte Listen von Wortformen jedoch keine Fälle von Homonymen erfassen. Wortformen werden auch als *tokens* bezeichnet. Unter dem Begriff *tokenizing* versteht man also die Erfassung der *tokens* eines Texts, wobei hier einige Standards, wie die Beachtung von Leerzeichen als Trennzeichen oder des Bindestrichs als Bestandteil des Lexems, gesetzt sein sollten.¹⁹⁷

Innerhalb eines Korpus wird nun die Anzahl der *tokens* gezählt und in einer sogenannten Frequenzliste angezeigt. Dabei werden die einzelnen Wortformen entweder alphabetisch oder nach Häufigkeit aufgelistet. Eine Angabe über die Größe des verwendeten Wortschatzes kann das Verhältnis zwischen *types* und *tokens* geben, die sogenannte *type-token-Relation* (TTR). Damit wird das Verhältnis zwischen allen unterschiedlichen Wortformen in einem Text (*types*) und allen im Text vorkommenden Lexemen insgesamt (*tokens*) bezeichnet. Je länger der Textausschnitt ist, den man wählt, desto höher ist die Anzahl der *tokens*, doch die Anzahl der *types* steigt mit zunehmender Textlänge immer langsamer aufgrund der Wiederholung von Lexemen.¹⁹⁸ Allerdings müssen hierbei zusätzlich die Begriffe *type* und ‚Lemma‘ unterschieden werden. Während ein *type* eine gewisse Wortform bezeichnet, vereint ein ‚Lemma‘ verschiedene Wortformen in sich.

Wortlisten können zum einen dazu genutzt werden, um sich einen Überblick über die häufigsten Lexeme einer Sprache zu verschaffen, zum anderen ermöglicht ein Vergleich von Wortlisten aus unterschiedlichen Korpora die Herausarbeitung von Spezialwortschatz. Vergleicht man beispielsweise die Frequenzliste eines medizinischen Korpus mit derjenigen eines Referenzkorpus, so können diejenigen Lexeme, welche im medizinischen Korpus häufig, im Referenzkorpus jedoch gar nicht oder selten vorkommen, als Fachwortschatz bezeichnet werden.¹⁹⁹

¹⁹⁷ Vgl. Gerstenberg (2013): op. cit., S. 91.

¹⁹⁸ Vgl. Gerstenberg (2013): op. cit., S. 92.

¹⁹⁹ Vgl. Scherer (2006): op. cit., S. 50.

Die Liste der 10.000 häufigsten Lexeme des *Corpus de Referencia del Español Actual* (CREA), eines großen Referenzkorpus der *Real Academia Española* (nähere Informationen zum CREA siehe Kapitel 2.1.1.1.), liefert folgende Ergebnisse, welche sowohl in Form der absoluten als auch der normalisierten Häufigkeit angegeben werden (zur Normalisierung, siehe Kapitel 1.4.7.):

Orden	Frec.absoluta	Frec.normalizada
1. de	9,999,518	65545.55
2. la	6,277,560	41148.59
3. que	4,681,839	30688.85
4. el	4,569,652	29953.48
5. en	4,234,281	27755.16
6. y	4,180,279	27401.19
7. a	3,260,939	21375.03
8. los	2,618,657	17164.95
9. se	2,022,514	13257.31
10. del	1,857,225	12173.87
11. las	1,686,741	11056.37
12. un	1,659,827	10879.95
13. por	1,561,904	10238.07
14. con	1,481,607	9711.74
15. no	1,465,503	9606.18

Abbildung 2: Beginn der Liste der 10.000 häufigsten Lexeme des CREA.²⁰⁰

Die frequentesten Lexeme sind demnach, wie auch zu erwarten ist, Funktionswörter wie Präpositionen, Artikel und Konjunktionen. Es muss jedoch beachtet werden, dass in dieser Häufigkeitsliste die einzelnen Wortformen aufgeführt werden und nicht die Lemmata. Ebenso verhält es sich auch mit den Artikeln, wie man an obiger Liste erkennen kann: Der bestimmte Artikel *la* erscheint im Singular über 6 Mio. Mal (Rang 2) und im Plural über 1,6 Mio. Mal (Rang 11).

1.4.2. Stichwortsuche – Suche nach Lexemen, Wortformen und Wortteilen

Die einfachste Art, sich mit einem Korpus zu beschäftigen, ist die Stichwortsuche. Dabei werden Lexeme, Wortformen oder Wortteile in die Suchmaschine eines Korpus eingegeben, und man lässt sich die Treffer anzeigen. Anhand dieser Möglichkeit kann die Existenz von Lexemen nachgeprüft werden, beispielsweise bei Neologismen. Dabei kann herausgefunden werden, ob ein Lexem wirklich neu ist oder ob es schon einmal in einem Text des Korpus erschienen ist. Auch die unterschiedlichen Verwendungsweisen und

²⁰⁰ Real Academia Española (ed.): *CREA. Listado de frecuencias. 10000 formas más frecuentes*. Online Publikation: http://corpus.rae.es/frec/10000_formas.TXT (24.02.2017).

gewöhnlichen Anschlüsse eines Lexems können anhand einer Stichwortsuche ermittelt werden.²⁰¹

Ein Problem für die Stichwortsuche stellen jedoch die Homographie und die Polysemie dar. Bei reinen Textkorpora sind keine Informationen zur Wortart oder zu Lemmata vorhanden, weshalb Suchabfragen anhand der Wortform alle Ergebnisse für die gleiche Form anzeigen.²⁰² Um diese Formen in einem nicht annotierten Korpus zu unterscheiden, bleibt daher lediglich der sprachliche Kontext, also die unmittelbare Textumgebung, wie sie in den Konkordanzen zu finden ist.

1.4.3. Konkordanzen und *Key Word in Context*

Lemnitzer/Zinsmeister definieren den Begriff ‚Konkordanz‘ wie folgt:

"Eine K. ist eine Sammlung von Kotexten eines bestimmten Schlüsselworts. Kotexte einer bestimmten Länge (von Buchstaben, Wörtern oder Sätzen) um ein Schlüsselwort herum werden aus einem Korpus extrahiert und meist mit dem Schlüsselwort im Zentrum angeordnet. Konkordanzen werden vor allem bei wortbezogenen Untersuchungen verwendet."²⁰³

In den gängigen Korpora werden die gefundenen Suchergebnisse meist als Konkordanzen dargestellt, das heißt, dass der gesuchte Begriff farblich hervorgehoben innerhalb seines Vorkommens im Text dargestellt wird. Diese Darstellung wird bisweilen auch als *Key Word in Context* (KWIC) bezeichnet.²⁰⁴

CONCORDANCIA

venido con él desde Madrid. Se llaman "Omar", una gata blanca, y su hijo, "Ulises", gato en gris perla, TARREGA (Urgell) Sábado, 22.30 horas. Teatro. "La gata damunt la teulada", por la Agrupació Cultural Re digrí para dedicarse a ladrarle a la gente, de la gata de Toña que es una belleza. Es tiempo de hablar l de honor. Desde las caraqueñas, pasando por 'La gata' Ibáñez, hasta la Virgen de Chiquinquirá, han si ae a los niños, sino también a los animalitos. La gata metió la pata con un gato tinaquero, del que se del que se enamoró perdidamente. Al principio, la gata se hizo la difícil ante los requerimientos amor e se mata es un feto. Si la hembra preñada es una gata, sus fetos son gatos; si se trata de una vaca, s d 1998 1998 10 504 R sierra de gata EN UN LUGAR DE LA RAYA El viajero que quiera con

Abbildung 3: Beispiel für Konkordanzen des Lexems *gata* im CREA.²⁰⁵

Konkordanzen dienen somit dazu, den sprachlichen Kontext einer Wortform zugänglich zu machen, ohne dabei jedoch eine Interpretation zu liefern. Dies ist die Aufgabe des Korpuslinguisten, der die gefundenen Daten analysiert und interpretiert, um zu einem Ergebnis zu gelangen.²⁰⁶

²⁰¹ Vgl. Scherer (2006): op. cit., S. 41-42.

²⁰² Vgl. Scherer (2006): op. cit., S. 43.

²⁰³ Lemnitzer/ Zinsmeister (2010): op. cit., S. 188.

²⁰⁴ Vgl. Gerstenberg (2013): op. cit., S. 93.

²⁰⁵ Real Academia Española (ed.): CREA. Buscar ‚gata‘. Online Publikation: <http://corpus.rae.es/cgi-bin/crpsrvEx.dll> (24.02.2017).

²⁰⁶ Vgl. Scherer (2006): op. cit., S. 45-46.

1.4.4. Suche mittels regulärer Ausdrücke

In einem Korpus kann jedoch nicht nur nach einzelnen Lexemen, Folgen von Lexemen oder Bestandteilen davon gesucht werden, sondern auch nach Wortmustern – anhand sogenannter regulärer Ausdrücke. Die Nutzung von regulären Ausdrücken beziehungsweise Platzhaltern und Operatoren ermöglicht die Suche nach bestimmten Textmustern in einer abgekürzten Form mittels sogenannter Metazeichen.²⁰⁷ Die folgende Übersicht zeigt die gängigsten regulären Ausdrücke:

Bezeichnung	Regulärer Ausdruck	Beispiel	Beschreibung
Alternation / Verknüpfung		(madre padre)	Findet alle Vorkommen von <i>madre</i> und <i>padre</i> . Die runden Klammern markieren den Bezugsbereich.
Gruppierung	()	(m p)adre (m p .)adre	Findet ebenfalls alle Vorkommen von <i>madre</i> und <i>padre</i> . Wenn alle Lexeme gesucht werden sollen, die auf <i>-adre</i> enden und insgesamt 5 Buchstaben haben, kann als erster Buchstabe ein Platzhalter gewählt werden, in diesem Fall ein Punkt. Nun wird beispielsweise auch <i>ladre</i> (von <i>ladrar</i>) gefunden.
Zeichenklasse	[]	[cfdl]ama	Findet alle Vorkommen von Lexemen wie <i>cama</i> , <i>fama</i> , <i>dama</i> und <i>lama</i> . Im Unterschied zur Alternation kann eine Zeichenklasse nur die Alternative für genau <u>ein</u> Zeichen auflisten.
Negierte Zeichenklasse	[^]	ca[^d]a	Findet alle vierstelligen Lexeme, die mit <i>ca-</i> beginnen und auf <i>-a</i> enden, aber kein <i>-d-</i> beinhalten: <i>casa</i> , <i>cara</i> , <i>caja</i> , <i>cama</i> , <i>capa</i> , <i>caza</i> , <i>caña</i> , aber nicht <i>cada</i> .
Platzhalter / Wildcard	.	c.ma	Findet beispielsweise Lexeme wie <i>cama</i> , <i>coma</i> , <i>cima</i> und <i>cuma</i> . Der Punkt steht für genau ein weiteres Zeichen.

²⁰⁷ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 89.

Operator für optionales Auftreten	?	ob?scuro	Das vorangehende Zeichen soll kein- oder einmal vorkommen. Findet die Formen <i>obscur</i> und <i>oscur</i> .
Operator für ein- oder mehrfaches Auftreten	+	ho+la	Findet die Formen <i>hola</i> , <i>hoola</i> , <i>hoolola</i> , <i>hooooola</i> , etc.
Operator für kein- oder einmaliges oder beliebig häufiges Auftreten	*	.*limp.*	Findet Formen wie <i>limpio</i> und <i>alimpiar</i> . Der Punkt als Platzhalter kann, muss aber nicht, ein oder mehrere Male ausgefüllt sein.
Umgekehrte Referenz / Maskierungszeichen	\	\.	Findet alle Vorkommen des Punktzeichens. Das nachfolgende Zeichen verliert seine Metabedeutung (in diesem Fall als Platzhalter).

Abbildung 4: Übersicht über die gängigsten regulären Ausdrücke.²⁰⁸

Die hier aufgelisteten regulären Ausdrücke stellen einen gewissen Standard dar, werden aber nicht in allen Programmen genauso verwendet. Die Operatoren *?*, *+* und *** werden beispielsweise für gewöhnlich mit dem Punkt als Platzhalter kombiniert, wenn ein beliebiges Zeichen gesucht werden soll (siehe vorletztes Beispiel in Abbildung 4). In einigen Programmen oder Korpora können diese Operatoren aber auch ohne den Punkt verwendet werden. In diesem Fall würde der vorletzte Suchausdruck **limp** lauten, aber dieselben Ergebnisse hervorbringen.

1.4.5. Annotation von Korpora: *tagging* und *parsing*

Einige Korpora bestehen jedoch nicht nur aus Primärdaten sowie Metadaten zu deren Entstehungskontext und –bedingungen, sondern auch aus Annotationen.²⁰⁹ Für deskriptive und textstrukturierende Annotationen gibt es die Empfehlungen der *Text-Encoding Initiative* (TEI). Das TEI-Konsortium entwickelt gemeinsam Standards für die Repräsentation von digitalen Texten und hält diese aufrecht. Seine Hauptaufgabe besteht in der Erstellung von Richtlinien zur Kodierung von maschinenlesbaren Texten. Diese Richtlinien, die hauptsächlich für die Geistes-, Sozial- und Sprachwissenschaften

²⁰⁸ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 90-92.

²⁰⁹ Vgl. Becker (2013): op. cit., S. 193-194.

bestimmt sind, finden seit dem Jahr 1994 weite Anwendung, vor allem in Büchereien und Museen sowie unter Verlegern und Wissenschaftlern.²¹⁰ Die TEI verwendet dafür die Auszeichnungssprache *Extensible Markup Language* (XML), welche eine vereinfachte Variante der *Standard Generalized Markup Language* (SGML) darstellt. Dabei werden die Informationen in spitzen Klammern angegeben, wobei zu Beginn eines Textbestandteils ausschließlich spitze Klammern verwendet werden (< >) und zum Ende spitze Klammern mit einem Schrägstrich (</ >), um eben dieses Ende zu markieren. Zudem besteht ein Vorteil der Arbeit mit XML darin, dass sie in unterschiedlichen Programmen verwendet und analysiert werden kann.²¹¹

Darüber hinaus gibt es die weitaus aufwändigeren, analytischen Annotationen. Dabei unterscheidet man zwischen morphosyntaktischen, syntaktischen, semantischen, pragmatischen und phonetischen Annotationen.

Die morphosyntaktische Annotation ist dabei die am häufigsten verwendete und bekannteste. Man bezeichnet sie auch als *grammatical tagging* oder *part-of-speech tagging* (kurz: *POS-tagging*) nach der englischen Bezeichnung für ‚Wortart‘. Bei dieser Art der Annotation wird jedes Lexem mit einem Label (engl. *tag*) versehen, welches dessen grammatikalische Klasse angibt. Hierdurch können Homographen disambiguiert werden. Die Summe aller verwendeten *tags* wird dabei als *tagset* bezeichnet und umfasst gewöhnlich 50 bis 150 solcher Labels.²¹²

Das *tagging*, also die Aufbereitung eines Korpus mit morphosyntaktischen Tags, ist ein aufwändiges und zeitintensives Unterfangen. Aus diesem Grund sind nicht alle Korpora getaggt, obgleich mit einem getaggtten Korpus viel besser und genauer gearbeitet werden kann. Um den Prozess zu vereinfachen, gibt es automatische *tagger*, also Computerprogramme, welche die einzelnen Lexeme mit Tags versehen. Diese computerlinguistischen Werkzeuge zeichnen sich jedoch durch eine hohe Anzahl an Fehlern aus, weshalb eine automatische Zuweisung allein meist nicht zufriedenstellend ist. Daher sollte die Annotation durch ein manuelles, von Linguisten durchgeführtes *tagging* komplettiert werden.²¹³

²¹⁰ Vgl. Text Encoding Initiative Consortium (ed.2016): *TEI: Text Encoding Initiative*. Online Publikation: <http://www.tei-c.org/index.xml> (24.11.2016).

²¹¹ Vgl. Gerstenberg (2013): op. cit., S. 89-90.

²¹² Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 65-66.

²¹³ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 63.

Neben dem Wortarten-Tagging gibt es die Annotation von Flexionsmorphologie, die sogenannte *Lemmatisierung*. Hier wird jedes *token* analysiert und auf das jeweilige Lemma zurückgeführt. Die Flexionsmorphologie umfasst dabei Kategorien wie Kasus, Genus, Numerus, Person, Tempus und Modus.²¹⁴

Ein Beispiel für ein lemmatisiertes und nach Wortarten getaggttes Korpus ist das *Corpus del Español (Género/Histórico)* von Mark Davies, dessen Vorgehensweise nun aus Gründen der Anschaulichkeit exemplarisch dargestellt werden soll (siehe auch Kapitel 2.1.1.3. und 2.2.1.3.). In diesem lemmatisierten Korpus werden alle Wortformen unter dem jeweiligen Lemma aufgelistet. Ambige Formen, die zu mehr als einem Lemma oder zu mehreren Wortarten gehören, werden durch den *tagger* disambiguiert und dem jeweiligen Lemma oder der jeweiligen Wortart zugeteilt.²¹⁵ Bei ambigen Wortformen wie *trabajo* taggten die Autoren des Korpus das Lexem in den Fällen, in denen *trabajo* mit vorangehendem bestimmtem Artikel *el* auftrat, als Singular des maskulinen Substantivs *trabajo*. Bei gemeinsamem Vorkommen mit vorangehendem Subjektpronomen *yo* wurde die Wortform als erste Person Singular Präsens des Verbs *trabajo* getaggt.²¹⁶

Allerdings gibt es auch Fälle, in denen die Unterscheidung weitaus schwieriger ist, wofür die Autoren probabilistische Informationen nutzten. Ein Beispiel für eine sehr schwer zu bestimmende Klasse von Lexemen ist das *participio perfecto*. Das regelbasierte Einordnungssystem des Korpus sucht bei solchen Fällen nach einer vorangehenden Form von *haber*, wie *he escrito*. In Fällen ohne *haber* wie *periódico escrito* kann *escrito* einerseits das *participio perfecto* des Verbs *escribir* sein, wie bei *leí el periódico escrito ayer*, im Sinne von ‘ich las die Zeitung, die gestern geschrieben worden war’, andererseits kann es aber auch mit adjektivischem Sinn gelesen werden, im Sinne von ‘die geschriebene Zeitung’ im Gegensatz zur elektronischen. In diesen Fällen analysierten die Autoren die Gesamtzahl der Fälle, in denen das *participio perfecto* auf das Verb *ser* folgt, was auf eine passive, verbale Lesart schließen lässt, und diejenigen Fälle, in denen es auf *estar* folgt, was auf eine resultative, adjektivische Lesart schließen

²¹⁴ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 70.

²¹⁵ Vgl. Davies, Mark (2005): „Vocabulary Range and Text Coverage: Insights from the Forthcoming *Routledge Frequency Dictionary of Spanish*“. In: Eddington, D. (Hrsg.) (2005): *Selected Proceedings of the 7th Hispanic Linguistics Symposium*. Somerville, MA: Cascadilla Proceedings Project, S. 108.

²¹⁶ Vgl. Davies, Mark (2006): *A Frequency Dictionary of Spanish. Core Vocabulary for Learners*. New York: Routledge, S. 4.

lässt. Falls ein bestimmtes *participio perfecto* häufiger mit *ser* auftrat, wurden diese Kombinationen als Passiv/Verb markiert.²¹⁷

Eine weitere sehr häufig verwendete Form ist die syntaktische Annotation. Korpora mit syntaktischer Annotation werden auch *Baumbanken* genannt, nach dem englischen *treebank*. Sätze werden hier in einer graphischen Struktur, als Baumdiagramm oder Klammerstruktur, dargestellt.²¹⁸ Die Markierung der *tokens* mit syntaktischen Kategorien wird als *parsing* bezeichnet. Hierfür wird beispielsweise der Teilsatz *Cada día contemplaba el cielo a través de la ventana* mit folgenden syntaktischen Kategorien versehen: *Cada día* [func="cc"] *contemplaba* [func="subj"] *el cielo* [func="cd"] *a través de la ventana* [func="cc"], wobei "cc" für *complemento circunstancial* steht, "subj" für *sujeto* und "cd" für *complemento directo*.²¹⁹ Für diesen Analysevorgang gibt es ebenso automatische Programme, die sogenannten *parsers*. Grammatisch-syntaktisch annotierte Texte gelten als *geparst*.²²⁰

Eine weitere Möglichkeit ist die semantische Annotation. Diese kann entweder auf der Wortebene stattfinden oder darüber hinausgehen, in Form von semantischen *frames*. Auf Wortebene ist allein die Abgrenzung der Eigennamen von gewöhnlichen Substantiven eine Art semantische Annotation. Außerdem ist es möglich, verschiedene Lesarten eines Lexems mittels *tags* zu markieren, beispielsweise die unterschiedlichen Bedeutungen des Lexems *contar* als ‚erzählen‘, ‚zählen‘ oder ‚mit jemandem rechnen‘. Die Anlage von *frames* innerhalb eines Korpus ist komplexer und findet sich daher seltener.²²¹

Zudem findet sich in manchen Korpora eine Annotation auf Lautebene. Hier können Merkmale der Aussprache (phonetische Annotation) sowie der Betonung und Intonation (prosodische Annotation) notiert werden, wobei festzuhalten ist, dass phonetisch oder prosodisch annotierte Korpora selten sind.²²²

Schließlich gibt es noch die Fehlerannotation. Diese Art der Annotation wird bei Lernerkorpora, welche auch Fehlerkorpora genannt werden, angewandt, um Fehler auf

²¹⁷ Vgl. Davies (2006): op. cit., S. 4.

²¹⁸ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 75.

²¹⁹ Der als Beispiel angegebene Teilsatz sowie die syntaktischen Funktionen wurden dem syntaktisch annotierten Korpus AnCorra entnommen (siehe Kapitel 2.5.2.3.). Vgl. Grupo de investigación CLiC (ed.): *AnCorra. Búsquedas*. Online Publikation: <http://clic.ub.edu/corpus/es/búsquedas> (16.06.2019).

²²⁰ Vgl. Kabatek/ Pusch (2011): op. cit., S. 213.

²²¹ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 85.

²²² Vgl. Scherer (2006): op. cit., S. 22.

den verschiedenen Ebenen, also beispielsweise Fehler der Aussprache, Orthographie, Morphologie, Wortstellung, der Kongruenz zwischen Wortformen oder bei der Wahl von Tempus und Modus festzuhalten. Die Lernerkorpora enthalten Texte von Personen, die eine Fremdsprache erlernen und dabei Fehler machen. Die in diesen Korpora gesammelten Daten können für den Sprachunterricht oder die Spracherwerbsforschung herangezogen werden.²²³

Für die eigene Erstellung von Annotationen gibt es im Internet frei verfügbare Annotationsprogramme wie CLaRK²²⁴ (entwickelt an den Universitäten Tübingen und Sofia), WordFreak²²⁵ und EXMARaLDA²²⁶ (entwickelt an der Universität Hamburg). Vor allem Letzteres eignet sich besonders für multimodale Annotationen.²²⁷

Während derzeit fast alle Korpora der romanischen Sprachen deskriptiv annotiert sind (vor allem nach TEI-Empfehlung), bedauern Kabatek/Pusch, dass es weiterhin nur wenige analytisch annotierte Korpora gibt, da deren Erstellung mit einem enormen Aufwand verbunden ist.²²⁸ Wie im zweiten Kapitel zu sehen sein wird, existieren dennoch einige analytisch annotierte Korpora der spanischen Sprache.

1.4.6. Abfragetools

Ein annotiertes Korpus an sich kann der Benutzer allerdings noch nicht verwenden, da dieses erst abfragbar gestaltet werden muss, um nach bestimmten Begriffen suchen oder gewisse Fragestellungen untersuchen zu können. Hierfür wird das Korpus in ein Abfragetool, meist ein Konkordanzprogramm, geladen.²²⁹ Die großen, im Internet zugänglichen Korpora sind alle über ein derartiges Abfragetool nutzbar, doch für die eigene Erstellung eines Korpus seien hier ein paar Programme genannt: *WordSmith*²³⁰,

²²³ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 86-87.

²²⁴ Vgl. Simov, Kiril/ Osenova, Petya: *CLaRK – an XML Based System for Corpora Development*. Online Publikation: <http://www.bultreebank.org/clark/index.html> (24.11.2016).

²²⁵ Vgl. tsmorton (ed.2004): *WordFreak*. Online Publikation: <http://wordfreak.sourceforge.net> (24.11.2016).

²²⁶ Vgl. Hamburger Zentrum für Sprachkorpora (ed.): *EXMARaLDA*. Online Publikation: <http://exmaralda.org/de/> (24.11.2016).

²²⁷ Vgl. Becker (2013): op. cit., S. 194.

²²⁸ Vgl. Kabatek/ Pusch (2011): op. cit., S. 214.

²²⁹ Vgl. Becker (2013): op. cit., S. 195-196.

²³⁰ Vgl. Lexical Analysis Software Ltd (ed.): *WordSmithTools*. Online Publikation: <http://www.lexically.net/wordsmith/> (21.02.2017).

*MonoConc*²³¹, *AntConc*²³², *Xaira*²³³ oder EXAKT²³⁴. Mittels dieser Konkordanzprogramme kann nach einzelnen Lexemen gesucht werden oder auch nach Wortmustern anhand regulärer Ausdrücke.

Als Ergebnis dieser Programme werden in der Regel diejenigen Sätze, welche den Suchbegriff enthalten, in ihrem sprachlichen Kontext, also als *Key Word in Context* (KWIC) angezeigt. Derartige Abfragetools erstellen zudem die bereits erwähnten Frequenzlisten (siehe Kapitel 1.4.1.), sie geben Kollokationen aus (siehe Kapitel 1.4.8.) und es kann eine *Keyword*-Analyse durchgeführt werden.²³⁵

1.4.7. Vergleichbarkeit von Daten: Normalisierung

Sofern ein Nutzer nun Daten von verschiedenen Korpora vorliegen hat und diese einander gegenüberstellen möchte, stellt sich ihm zwangsläufig die Frage, auf welche Weise ein Vergleich sinnvoll ist. Manchmal möchte man beispielsweise die Häufigkeit bestimmter Suchbegriffe in verschiedenen Korpora miteinander vergleichen. Wenn man die Frequenzen von drei spanischen Lexemen (*instituto*, *papel* und *carretera*) in drei der größten spanischen Korpora (darunter zwei Referenzkorpora) miteinander vergleicht, ergibt sich zunächst die folgende Übersicht:

Suchbegriff	CdE:Nuevo	CREA	CORPES
<i>instituto</i>	167.560	2.142	5.838
<i>digital</i>	143.644	3.121	10.283
<i>carretera</i>	66.795	8.483	11.409

Abbildung 5: Ergebnisse der Stichwortsuche in den Korpora CdE:Nuevo, CREA und CORPES (absolut).²³⁶

Zum einen ist zu erkennen, dass alle Lexeme am häufigsten im CdE:Nuevo (*Corpus del Español: Nuevo* – siehe Kapitel 2.6.1.) erscheinen. Dies könnte man anhand der Zusammensetzung der Korpora erklären: Während das bereits erwähnte CREA (*Corpus de Referencia del Español Actual* – siehe Kapitel 2.1.1.1.) und das CORPES (*Corpus del Español del Siglo XXI* – siehe Kapitel 2.1.1.2.) Referenzkorpora sind und die ganze

²³¹ Vgl. Athelstan (ed.): *MonoConc*. Online Publikation: <http://www.athel.com/mono.html> (21.02.2017).

²³² Vgl. Anthony, Laurence (2018): *AntConc (Version 3.5.7) [Macintosh OS X]*. Tokyo, Japan: Waseda University. Online Publikation: <http://www.antlab.sci.waseda.ac.jp/> (14.05.2018).

²³³ Vgl. Sourceforge (ed.): *Xaira*. Online Publikation: <http://xaira.sourceforge.net> (21.02.2017).

²³⁴ Vgl. Schmidt, Thomas/ Wörner, Kai/ Lehmberg, Timm/ Hedeland, Hanna: *EXACT*. Online Publikation: <http://exmaralda.org/de/exakt-de/> (21.02.2017).

²³⁵ Vgl. Becker (2013): op. cit., S. 196.

²³⁶ Hierbei gilt es zu beachten, dass in allen Korpora nach der jeweiligen Wortform gesucht wird und nicht nach dem Lemma, um einheitliche Ergebnisse zu erhalten.

Bandbreite geschriebener und gesprochener Texte des Spanischen integrieren, handelt es sich bei den Quellen des CdE:Nuevo hauptsächlich um aktuelle Texte von Internetseiten. Somit könnte gefolgert werden, dass diese drei Lexeme schlichtweg sehr häufig im Internet erscheinen. Es sollte allerdings auch immer die Größe des jeweiligen Korpus beachtet werden. Wenn man den Umfang der drei Korpora betrachtet, so ist das CdE:Nuevo mit zwei Milliarden Lexemen derzeit das größte, gefolgt vom CORPES mit 225 Mio. Lexemen und das CREA ist das verhältnismäßig kleinste Korpus mit 160 Mio. Lexemen. Möchte man nun Vergleiche zwischen den Ergebnissen der Korpora anstellen, so müssen die Ergebnisse auf eine bestimmte Anzahl von Lexemen pro Text gebracht werden. Dieser Vorgang nennt sich *Normalisierung*. Die Normalisierung kann sich auf 10.000 Lexeme oder auf eine Mio. beziehen, sie sollte sich jedoch an der typischen Textlänge im Korpus orientieren.²³⁷

Werden die Ergebnisse auf eine Mio. Lexeme normalisiert, so ergibt sich folgende Übersicht:

Suchbegriff	CdE:Nuevo	CREA	CORPES
<i>instituto</i>	83	13	25
<i>digital</i>	71	19	45
<i>carretera</i>	33	53	50

Abbildung 6: Ergebnisse der Stichwortsuche in den Korpora CdE:Nuevo, CREA und CORPES (normalisiert je eine Mio. Lexeme).

Vergleicht man die normalisierten Ergebnisse, so fällt auf, dass die Lexeme *instituto* und *digital* zwar weiterhin im CdE:Nuevo am häufigsten sind, das Lexem *carretera* allerdings häufiger im CORPES und im CREA erscheint. Anhand dieses Ergebnisses könnte man schließen, dass *carretera* tatsächlich in Internettexten weniger häufig vorkommt als im Durchschnitt spanischer Texte allgemein, so wie diese in den Korpora CORPES und CREA zu finden sind. Zudem ist zu erkennen, dass das Lexem *digital* im CORPES häufiger vorkommt als im CREA, was damit erklärt werden könnte, dass es sich beim CORPES, das aus Texten des 21. Jahrhunderts besteht, um aktuellere Texte handelt als im Fall des CREA.

Generell ist beim Vergleich von Daten aus verschiedenen Korpora immer auf die unterschiedlichen Korpustypen, die Korpusgröße sowie die Gewichtung der Korpora zu

²³⁷ Vgl. Scherer (2006): op. cit., S. 38-39.

achten. Vergleiche zwischen Varietätenkorpora und Referenzkorpora oder zwischen Korpora unterschiedlicher diachroner Abschnitte sowie zwischen Lernerkorpora und vergleichbaren Spezialkorpora von Muttersprachlern erscheinen beispielsweise grundsätzlich als zielführend, ein Vergleich zwischen einem diachronen Korpus und einem Lernerkorpus jedoch eher nicht.

1.4.8. Kollokationen und statistische Verfahren für deren Ermittlung

Korpora erfreuen sich in der Sprachwissenschaft allerdings nicht nur wegen der Analyse der Frequenz einzelner Lexeme großer Beliebtheit, sondern auch und vor allem wegen ihrer Möglichkeiten zur Ermittlung von Kollokationen (zum Begriff der Kollokation und deren Abgrenzung von Phraseologismen siehe Kapitel 4.1.1. und 4.1.2.).

Da bei der Ermittlung von Kollokationen nicht die bloße Häufigkeit ausschlaggebend ist, sondern unter anderem auch kombinatorische Restriktionen, aufgrund derer ein Bestandteil der Kollokation teilweise kaum außerhalb dieser Kombination verwendet wird, spielt der Vergleich der Frequenz der beiden Einheiten an sich mit der Häufigkeit des gemeinsamen Vorkommens als Kollokation eine wichtige Rolle. Ist dieses Vorkommen häufiger, als es allein durch die zufällige Verteilung der Lexeme zu erwarten wäre, so kann man davon ausgehen, dass es sich um eine Kollokation handelt. Zur Ermittlung dieses Verhältnisses gibt es drei verschiedene Verfahren: *Mutual Information*, *T-Score* und *LL-Simple*.

Mittels des statistischen Verfahrens *Mutual Information* (MI) wird die Häufigkeit des gemeinsamen Vorkommens linguistischer Einheiten innerhalb eines Sprachausschnitts mit der erwarteten Frequenz aufgrund der individuellen Häufigkeiten der Elemente in diesem Sprachausschnitt verglichen. Sein Wert drückt den Grad der Abweichung zwischen der realen und der erwarteten Frequenz aus und gibt den Grad der Konditionierung an, den eine Einheit über seinen linguistischen Kontext ausübt. Der Wert wird als Index für die Assoziationskraft zwischen den Einheiten gesehen. Das Verfahren *T-Score* zieht neben der Divergenz zwischen der realen und der erwarteten Häufigkeit einer Kombination auch die Anzahl der Belege in Betracht, welche der Sprachausschnitt zur Berechnung vorbringt, um diese Divergenz zu begründen. Sein Wert wird daher als Index des Vertrauens gesehen, welcher der kalkulierten Assoziationskraft der Einheiten beizumessen ist. Theoretisch wird davon ausgegangen, dass er eine Verteilung der in den Texten abwesenden Daten veranschlagt. In der Praxis stellt sich jedoch heraus, dass er

hochfrequente Kombinationen bevorzugt, weshalb er speziell für die Erkennung besonders konsistenter Muster verwendet wird. Das Verfahren *LL-Simple* (*Log Likelihood*) vergleicht wie die MI die reale und die erwartete Häufigkeit, es zieht allerdings auch die Menge der verfügbaren Belege in Betracht, und sein Wert ist in Übereinstimmung mit einer normalisierten Skala zu interpretieren. Obwohl diese normalisierte Skala es ermöglicht, Fehlertoleranzen anzugeben, überschreiten viele Paare mit hoher Frequenz, aber schwacher Assoziation eine sehr wenig restriktive Toleranzschwelle. Somit bevorzugen *LL-Simple* und vor allem *T-Score* Kombinationen hoher Frequenz, wodurch auf den ersten Positionen teilweise Kombinationen von geringem Interesse oder ohne eindeutige lexikalische Assoziation stehen.²³⁸ Im Verlauf dieser Arbeit, insbesondere in Kapitel 4.1.1., wird zu sehen sein, dass auch in einigen spanischen Korpora Kollokationen mittels eines oder mehrerer dieser statistischen Verfahren ermittelt werden können.

Wie in diesem ersten Kapitel zu sehen war, können Korpora auch für den Sprachunterricht eine Rolle spielen, da sie Konstruktionen auf der Skala der Akzeptabilität verorten und damit Rückschlüsse auf die Sprachnorm zulassen. Eine Grundvoraussetzung dafür ist die Repräsentativität des Korpus, welche bei der Erstellung desselben auf jeden Fall angestrebt werden sollte, denn nur ein Korpus, das für einen Sprachausschnitt repräsentativ ist, kann verlässliche Daten für diesen liefern.

Die Frage nach dem Stellenwert der Korpuslinguistik als Methode oder Disziplin kann nicht eindeutig geklärt werden, denn sie ist neben der fachwissenschaftlichen Diskussion auch von der wissenschaftstheoretischen Ausrichtung des jeweiligen Forschers und der Art der Korpusnutzung für die eigene Forschung abhängig. Obgleich die Korpuslinguistik heute eher als Disziplin zu sehen ist, spielt diese Unterscheidung für die vorliegende Arbeit ohnehin eine untergeordnete Rolle, denn es soll geklärt werden, inwiefern Korpora sinnvoll im und für den Fremdsprachenunterricht genutzt werden können. Somit ist der Status der Korpuslinguistik an sich für die Klärung dieses Anliegens zweitrangig. In diesem Sinne werden in dieser Arbeit, wie bereits angegeben, alle

²³⁸ Vgl. Real Academia Española (ed.): *CORPES. Manual de uso*. Online Publikation: <http://web.frl.es/CORPES/org/publico/pages/ayuda/manual.view> (20.10.2017), S. 17-18.

1. Korpora und Korpuslinguistik

Studien, die anhand eines Korpus durchgeführt wurden, einheitlich als ‚korpusbasiert‘ bezeichnet.

Des Weiteren konnte in diesem Kapitel gesehen werden, dass Korpora einige Verfahren und Werkzeuge bieten, die auch für den Fremdsprachenunterricht interessant erscheinen. Hierbei sticht vor allem die Option der Suche nach Kollokationen ins Auge, aber auch die Auflistung von Konkordanzen mittels einer einfachen Stichwortsuche oder die Ermittlung der Häufigkeit eines Ausdrucks könnten für den Unterricht genutzt werden, wie in Kapitel 3 und insbesondere Kapitel 4 zu sehen sein wird. Um nun mehr über die verschiedenen Arten von Korpora herauszufinden, anhand derer die eben vorgestellten Funktionen ausgeführt werden können, werden die Korpora der spanischen Sprache im Folgenden in Kategorien unterteilt und genauer erläutert.

2. Kategorisierung der Korpora

In der folgenden Einteilung der Korpora in verschiedene Subgattungen wird weitgehend nach den Kriterien von Lemnitzer/Zinsmeister vorgegangen. Es handelt sich bei der Auflistung in diesem Kapitel um eine Kategorisierung der unterschiedlichen Korpusarten mit direkt daran anschließenden Informationen zu entsprechenden spanischen Beispielen, wobei kein Anspruch auf Vollständigkeit erhoben werden soll. Es sollen lediglich einige Korpora aus dem spanischsprachigen Raum gezeigt werden, welche auch im Unterricht verwendet werden können, wobei die Möglichkeiten zur Nutzung derselben erst im vierten Kapitel beschrieben werden. Lemnitzer/Zinsmeister setzen für die Kategorisierung von Korpora die Kriterien Funktionalität, Sprachenauswahl, Medium, Größe, Annotation, Persistenz, Sprachbezug und Verfügbarkeit ein. Ebenso wird in diesem Kapitel verfahren.²³⁹

Bislang konnte sich der Nutzer über die verschiedenen Arten von Korpora und spanische Beispiele dafür lediglich in Werken zur allgemeinen Einführung in die spanische Sprachwissenschaft wie Becker²⁴⁰ oder Kabatek/Pusch²⁴¹ informieren. Pusch veröffentlichte zusätzlich bereits mehrmals reine Auflistungen von Korpora gesprochener romanischer Sprachen im Internet. Viele der dort erwähnten Korpora sind jedoch bereits veraltet und lediglich als Buch, Diskette oder CD-ROM verfügbar.²⁴² Bei Koch/Oesterreicher findet sich ebenfalls eine Auflistung der spanischen Korpora gesprochener Sprache.²⁴³ Außerdem liegen einige Online-Datenbanken vor, wie die ‚Textdatenbank Romanistik‘ der Freien Universität Berlin²⁴⁴, oder Blogs, wie die ‚Linguistische Werkstatt‘ der Universität Bremen²⁴⁵. Zudem gibt es auch Internetseiten spanischer Universitäten, welche über verfügbare Korpora informieren, so zum Beispiel

²³⁹ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 102-107.

²⁴⁰ Vgl. Becker (2013): op. cit., S. 201-204.

²⁴¹ Vgl. Kabatek/ Pusch (2011): op. cit., S. 215-218.

²⁴² Vgl. Pusch, Claus D. (2014): *A survey of spoken language corpora in Romance*. Online Publikation: http://www.romanistik.uni-freiburg.de/pusch/Download/publikationen/2002_Survey_corpora.pdf (07.03.2017), S. 5-9.

²⁴³ Vgl. Koch/ Oesterreicher (2011): op. cit., S. 37-39.

²⁴⁴ Vgl. Pascher, Franziska (2015): *Spanische Korpora und Textdatenbanken*. Online Publikation: <http://wikis.fu-berlin.de/display/tdr/Spanische+Korpora+und+Textdatenbanken> (06.03.2017).

²⁴⁵ Vgl. Uni Bremen Blogs (ed.): *Linguistische Werkstatt. Spanische Korpora*. Online Publikation: <https://blogs.uni-bremen.de/lingwerk/spanische-korpora/> (06.03.2017).

2. Kategorisierung der Korpora

die Übersicht von Joaquim Llisterri von der *Universitat Autònoma de Barcelona*²⁴⁶, sowie eine, wenn auch relativ spärliche Übersicht der *Linguist List*.²⁴⁷ Diese Werke und Internetseiten zeigen jedoch immer nur einen kleinen Ausschnitt der Vielfalt der spanischen Korpora. Zudem werden sie dort meist nur namentlich erwähnt und weder beschrieben noch korpusstypologisch eingeordnet. Nach einer Auflistung der verschiedenen Korpuskategorien inklusive einer dementsprechenden Einteilung der gängigsten Korpora der spanischen Sprache hat man bisher vergeblich gesucht, weshalb die vorliegende Kategorisierung durchaus einen gewissen Neuheitswert vorzuweisen hat. Alle in diesem Kapitel beschriebenen Korpora finden sich zudem in der alphabetischen Übersicht über die Korpora der spanischen Sprache in Anhang 4. Es handelt sich dabei nicht nur um eine alphabetisch geordnete Korpusliste, sondern es werden zudem die wichtigsten Informationen zu jedem Korpus auf einen Blick ersichtlich.

2.1. Sprachbezug

Bei diesem Kriterium wird zwischen *Referenzkorpora* und *Spezialkorpora* unterschieden. Erstere zeichnen sich dadurch aus, dass sie die Eigenschaften der repräsentierten Sprache zur jeweiligen Zeit möglichst gut abdecken. Während ein Referenzkorpus ausgewogen sein, das heißt, der Heterogenität der Sprache gerecht werden sollte, bezieht sich ein Spezialkorpus auf eine spezifische Varietät, wobei die zu untersuchenden Phänomene dieser Varietät sichtbar werden, wenn man deren Verteilung im Spezialkorpus und im Referenzkorpus vergleicht. Das Referenzkorpus dient hierbei als Hintergrund- oder Kontrollkorpus.²⁴⁸

Ein weiterer Begriff, der hier zu erwähnen ist, obgleich er keinen eigenständigen Wert in der Korpusstypologie darstellt, ist das *opportunistische Korpus*. Bei der Korpuserstellung stellt die Beschaffung von Texten vor allem bei wohldefinierten Designkriterien eine enorme Hürde dar. Aus diesem Grund werden manche Texte schlichtweg deshalb in ein Korpus integriert, weil sie verfügbar sind und keinen Copyright-Beschränkungen unterliegen. Ein Korpus, das aus diesem Grund eben nicht

²⁴⁶ Vgl. Llisterri, Joaquim: *Corpus de lengua oral en español*. Online Publikation: http://liceu.uab.es/~joaquim/language_resources/spoken_res/Corp_leng_oral_esp.html (07.03.2017).

²⁴⁷ Vgl. The Linguist List (ed.): *Texts & Corpora related to Spanish*. Online Publikation: <http://linguistlist.org/sp/SearchWRListing-action.cfm?subclassid=5727&SearchType=SL&WRTypeID=1> (07.03.2017).

²⁴⁸ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 106.

repräsentativ, aber sehr groß ist, weil versucht wird, alle verfügbaren Texte zu integrieren, nennt man ein opportunistisches Korpus. Bei einem opportunistischen Korpus handelt es sich im Normalfall um ein Spezialkorpus, zum Beispiel um eine Sammlung von Artikeln mehrerer Jahrgänge einer Tageszeitung.²⁴⁹

2.1.1. Referenzkorpora

Die *Real Academia Española* definiert ein Referenzkorpus wie folgt:

“Un corpus de referencia es aquel que está diseñado para proporcionar información exhaustiva acerca de una lengua en un momento determinado de su historia y, por tanto, ha de ser lo suficientemente extenso para representar todas las variedades relevantes de la lengua en cuestión.”²⁵⁰

Ein Referenzkorpus muss also ausreichend groß sein, um alle relevanten Varietäten der Sprache zu einem gegebenen Zeitpunkt in der Geschichte derselben zu repräsentieren. Mit dieser Definition beschreibt die *Real Academia Española* ihr *Corpus de Referencia del Español Actual* (CREA).

Laut Lemnitzer/Zinsmeister werden Referenzkorpora für eine Vielzahl linguistischer Untersuchungen verwendet. Ergebnisse, welche auf einem Referenzkorpus beruhen, können besser nachvollzogen und verglichen werden. Daher sollte ein solches Korpus bezüglich des abgebildeten Gegenstands einen hohen Grad der Abdeckung und strukturellen Ähnlichkeit aufweisen.²⁵¹

Teubert/Čermáková geben an, ein Referenzkorpus bestehe aus den Texten, die eine Sprachgemeinschaft für sich selbst als repräsentativ auswähle (siehe Kapitel 1.1.2.). Diese Auswahl ist allerdings keineswegs konstant, sondern sie ändert sich mit der Zeit. So sollte ein Referenzkorpus für gewöhnlich das Pensum an Literatur eines gebildeten Mitglieds der Mittelklasse widerspiegeln und auch mündliche Texte, die ein solches Mitglied der Gesellschaft normalerweise im Gespräch, zu förmlichen gesellschaftlichen Anlässen oder im Radio hört, enthalten.²⁵² Die Frage nach der Repräsentativität eines Korpus hängt allerdings auch immer von der beabsichtigten Nutzung, der Spezialisierung der Texte, der Anzahl der Sprachen sowie weiteren Kriterien ab. Ein Korpus mit einer

²⁴⁹ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 106-107.

²⁵⁰ Real Academia Española (ed.): *CREA Escrito y Oral*. Online Publikation: <http://www.rae.es/rae/gestores/gespub000019.nsf/voTodosporId/B104F9F0D0029604C1257164004032BE?OpenDocument&i=1> (19.05.2011).

²⁵¹ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 190.

²⁵² Vgl. Teubert/ Čermáková (2007): op. cit., S. 65-66.

2. Kategorisierung der Korpora

Mio. Lexemen über spanische Wetterberichte ist daher viel repräsentativer als ein Korpus spanischer Literatur dieser Größe.²⁵³ Ein Referenzkorpus einer Nationalsprache sollte laut Teubert/Čermáková heute zwischen 50 und 500 Mio. Lexeme oder sogar mehr umfassen.²⁵⁴

Referenzkorpora dienen einer Vielzahl an Verwendungsmöglichkeiten. Sie enthalten das Standardvokabular und sind die Hauptquelle des Korpuslinguisten, um mehr über die Bedeutung von Lexemen herauszufinden. Wenn sie groß genug sind, können die typischen Kontexte eines Lexems gesucht und Kollokationen gefunden, sowie Mehrdeutigkeit aufgedeckt werden. Referenzkorpora können zudem, wie bereits erwähnt, als Bezugspunkt für Spezialkorpora dienen.²⁵⁵

Für das Spanische gibt es derzeit zwei große Referenzkorpora, welche im Folgenden vorgestellt werden sollen: Das *Corpus de Referencia del Español Actual* (CREA) und das *Corpus del Español del Siglo XXI* (CORPES). Bei beiden handelt es sich um Korpora der *Real Academia Española*.

Auch das Subkorpus des aktuellen Spanisch, ein Bestandteil des ursprünglichen *Corpus del Español* (CdE:Orig), kann mit über 20 Mio. Lexemen durchaus als Referenzkorpus bezeichnet werden.²⁵⁶ Da dieses Teilkorpus ein Referenzkorpus darstellt, jedoch auch Bestandteil des diachronen Korpus CdE:Orig ist, wird das Korpus in Teilen sowohl im vorliegenden Abschnitt als auch im Unterkapitel ‚Funktionalität‘ beschrieben werden. Die Einordnung des CdE:Orig bereitete schon mehreren Linguisten Probleme, was auch daran zu erkennen ist, dass es bei Becker²⁵⁷ beispielsweise als ‚Korpus der spanischen Gegenwartssprache‘ beschrieben wird und bei Kabatek/Pusch²⁵⁸ als diachrones Korpus.

2.1.1.1. CREA – *Corpus de Referencia del Español Actual*

Das *Corpus de Referencia del Español Actual* (CREA) bezieht sich ausschließlich auf modernes Spanisch und wurde zusammen mit dem *Corpus Diacrónico del Español*

²⁵³ Vgl. Borja, Anabel (2008): „Corpora for Translators in Spain. The CDJ-GITRAD Corpus and the GENTT Project“. In: Andermann, G./ Rogers, M. (Hrsg.) (2008): *Incorporating Corpora. The linguist and the translator*. Clevedon [u.a.]: Multilingual Matters, S. 247.

²⁵⁴ Vgl. Teubert/ Čermáková (2007): op. cit., S. 67.

²⁵⁵ Vgl. Teubert/ Čermáková (2007): op. cit., S. 67-68.

²⁵⁶ Vgl. Davies (2006): op. cit., S. 2.

²⁵⁷ Vgl. Becker (2013): op. cit., S. 202.

²⁵⁸ Vgl. Kabatek/ Pusch (2011): op. cit., S. 216.

(CORDE – siehe Kapitel 2.2.1.1.) im Jahr 2001 von der *Real Academia Española* veröffentlicht.²⁵⁹ Es ist im Internet kostenlos und frei zugänglich (<http://corpus.rae.es/creanet.html>).²⁶⁰

Die *Real Academia Española* beschreibt dieses Korpus als Referenzkorpus und begründet ihre Zuordnung mit der großen Anzahl an Lexemen, die das CREA umfasst.²⁶¹ Der Aufbau des CREA entspricht der Intention der Autoren, den Sprachforschern sowie anderen Interessierten einen repräsentativen und ausgeglichenen Ausschnitt der für die gesamte Hispanophonie gültigen Standardvarietät zu bieten. Um größtmögliche Flexibilität bei der Ermittlung der Daten zu erlauben, ist das CREA in verschiedene Module unterteilt, so dass es möglich ist, sowohl im gesamten Korpus als auch nur in Texten mit bestimmten geographischen, thematischen oder zeitlichen Charakteristika zu suchen.²⁶²

Auf der Suche nach dem Gleichgewicht zwischen der größtmöglichen Anzahl an Formen und der Möglichkeit, das Korpus mittels Kodifikation und grammatikalischer Annotation zu bereichern, umfasste das CREA am Ende der zweiten Phase, also im Dezember des Jahrs 2000, 125 Mio. Formen.²⁶³ Seit Juni 2008 zählt es nun mehr als 160 Mio. Lexeme.²⁶⁴ Davon stammen 90% aus geschriebenen und 10% aus gesprochenen Quellen.²⁶⁵ Von den geschriebenen Texten entfallen 49% auf Bücher, 49% auf die Presse und die verbleibenden 2% auf verschiedene Texte, wie Broschüren, Prospekte, E-Mails und Weblogs. Die Transkriptionen der mündlichen Quellen stammen weitgehend aus Radio und Fernsehen.²⁶⁶

Die Texte wurden ursprünglich in vier verschiedenen Medien veröffentlicht – in Büchern, Zeitschriften, Zeitungen sowie als Transkriptionen mündlicher Texte. Des Weiteren sind sie auf die Themen Wissenschaft und Technologie, Geistes- und Sozialwissenschaften, Politik und Wirtschaft, Kunst, Gesundheit, Freizeit und

²⁵⁹ Vgl. Davies/ Face (2006): op. cit., S. 133.

²⁶⁰ Vgl. Real Academia Española (ed.): *CREA (Version 2015)*. Online Publikation: <http://corpus.rae.es/creanet.html> (19.05.2011).

²⁶¹ Vgl. Real Academia Española (ed.): *CREA Escrito y Oral*. Op. cit. (19.05.2011).

²⁶² Vgl. Real Academia Española (ed.): *CREA. Ayuda*. Online Publikation: http://corpus.rae.es/ayuda_c.htm (14.05.2011).

²⁶³ Vgl. Real Academia Española (ed.): *CREA. Ayuda*. Op. cit. (14.05.2011).

²⁶⁴ Vgl. Real Academia Española (ed.): *CREA Escrito y Oral*. Op. cit. (19.05.2011).

²⁶⁵ Vgl. Real Academia Española (ed.): *CREA. Ayuda*. Op. cit. (14.05.2011).

²⁶⁶ Vgl. Real Academia Española (ed.): *CREA Escrito y Oral*. Op. cit. (19.05.2011).

Alltagsleben sowie Erzählliteratur aufgeteilt.²⁶⁷

Geographisch gesehen, stammt die Hälfte der Texte aus Spanien und die andere Hälfte aus Lateinamerika. Bei den Texten aus Lateinamerika wurde versucht, mittels der Auswahl von Texten aus verschiedenen linguistischen Zonen die dortige Vielfalt widerzuspiegeln. Je nach der Bevölkerungszahl und dem kulturellen Gewicht der jeweiligen Sprachzone wurde aus den einzelnen hispanophonen Ländern beziehungsweise Zonen ein bestimmter Prozentsatz an Texten integriert.²⁶⁸ Es handelt sich dabei zumeist um solche aus den sprachlichen Zonen Mexiko (40%), der Anden (20%) und der Karibik (17%). Weiter umfasst das Korpus Texte aus der Rio de la Plata-Gegend (14%) und aus Chile (6%) sowie eine geringe Anzahl aus der zentralen Zone Lateinamerikas (3%).²⁶⁹

Ausgehend von der Annahme, dass die voraussichtliche Größe des Korpus die vorausgesetzte Varietät und Repräsentativität garantiert, hat man sich dafür ausgesprochen, im CREA in allen Fällen die kompletten Texte zu integrieren.²⁷⁰ Das CREA ist jedoch nur als Datenbank online zugänglich, der Gesamttext kann nicht angesehen oder heruntergeladen werden.²⁷¹

Durch die genaue Auflistung der geographischen, zeitlichen, thematischen und medialen Herkunft der Texte des CREA kann der Nutzer durchaus erkennen, dass es sich bei diesem Korpus tatsächlich um ein repräsentatives Korpus handelt. Das CREA kann also, wie von der *Real Academia Española* vorgeschlagen, als Referenzkorpus bezeichnet und auch als solches genutzt werden. Die baldige Integration weiterer neuer Texte aus den Jahren nach 2000 wäre äußerst wünschenswert, und es ist auch legitim, sich zu fragen, ob denn im Verhältnis nicht mehr Texte aus Lateinamerika stammen beziehungsweise ob es nicht generell mehr Texte gesprochener Sprache sein sollten, doch diese Entscheidungen müssen bei der Erstellung eines Korpus getroffen werden, und die Ersteller müssen zwischen derartigen Zuteilungen abwägen. Wie im kommenden Abschnitt zu sehen sein wird, ist das neue Korpus der *Real Academia Española*, das CORPES, ganz anders aufgebaut und legt damit sein Hauptaugenmerk auf andere

²⁶⁷ Vgl. Real Academia Española (ed.): *CREA. Ayuda*. Op. cit. (14.05.2011).

²⁶⁸ Vgl. Real Academia Española (ed.): *CREA. Ayuda*. Op. cit. (14.05.2011).

²⁶⁹ Vgl. Real Academia Española (ed.): *CREA. Estadísticas*. Online Publikation: <http://buscon.rae.es/nomina/SrvltGUIBusTextos?est=1> (14.05.2011).

²⁷⁰ Vgl. Real Academia Española (ed.): *CREA. Ayuda*. Op. cit. (14.05.2011).

²⁷¹ Vgl. Koch/ Oesterreicher (2011): op. cit., S. 38.

Kriterien.

Wie bereits erwähnt, ist seit Juni 2008 im Internet die neue Version des CREA frei zugänglich. Die neue Fassung ist um 3,5 Mio. Lexeme erweitert (damit derzeit 160 Mio. Lexeme), welche alle aus Texten aus der Zeit zwischen den Jahren 2000 und 2004 stammen. Die meisten dieser Texte sind aus der lateinamerikanischen Presse, wenn auch einige aus Büchern stammen. Den Erstellern ist demnach nicht entgangen, dass zunehmend neue Texte, vor allem aus Lateinamerika, in das CREA integriert werden sollten. Die neue Version enthält zusätzlich zu den Listen der häufigsten 1000, 5000 und 10000 Lexemen auch eine Liste mit der absoluten und der normalisierten Häufigkeit aller im CREA vorkommenden *tokens* (siehe Kapitel 1.4.1.). Des Weiteren wird bei der Erstellung der Wortlisten auf die Unterscheidung zwischen Groß- und Kleinschreibung verzichtet, sodass alle *tokens* zusammengezählt werden.²⁷²

Im November 2015 wurde die annotierte Version des CREA (0.1) veröffentlicht. In dieser Version wird die Suche nach Formen, Lemmata und grammatikalischen Kategorien ermöglicht (<http://web.frl.es/CREA/view/inicioExterno.view>).²⁷³ So kann in der neuen Version beispielsweise explizit nach *para* als Präposition gesucht werden, ohne dabei Verbformen von *parar* oder *parir* zu erhalten:

The screenshot shows the search interface of the annotated CREA version. At the top, there is a navigation bar with links: **Concordancias**, **Coapariciones**, **Configuración**, **Ayuda**, **Modo de cita**, and **Sugerencias**. Below this, there is a search form with the following fields: **Lema** (containing 'para'), **Forma** (containing 'para'), and **Clase de palabra** (containing 'preposición'). There are also buttons for **+** and **-** to expand or collapse the form. At the bottom of the form, there are three buttons: **Concordancia** (highlighted in red), **Estadística** (highlighted in red), and **Nueva consulta**.

Abbildung 7: CREA – Suchmaske der annotierten Version des CREA bei der Suche nach der Präposition *para*.²⁷⁴

²⁷² Vgl. Real Academia Española (ed.): *Banco de datos. Nueva version del CREA (textos de junio 2008)*. Online Publikation: <http://www.rae.es/rae/gestores/gespub000019.nsf/voTodosporId/E1E307BE0FBEE6FFC1257464003B0B63?OpenDocument> (20.05.2011).

²⁷³ Vgl. Real Academia Española (ed.): *Banco de datos. CREA*. Online Publikation: <http://www.rae.es/recursos/banco-de-datos/crea> (12.12.2016).

²⁷⁴ Real Academia Española (ed.): *CREA. Versión anotada*. Online Publikation: <http://web.frl.es/CREA/view/inicioExterno.view> (17.04.2018).

2. Kategorisierung der Korpora

CONCORDANCIA		Ordenar por:	Año ascendente	sin criterio
un conjunto de condiciones morfológicas y funcionales que son las más apropiadas	para	facilitar la fecundación y la implantación. En el ciclo femenino normal, durante		
unos días se alcanza una situación bioquímica y fisiológica óptima, la más adecuada	para	la implantación del óvulo fecundado y para su desarrollo hasta la fase terminal		
y fisiológica óptima, la más adecuada para la implantación del óvulo fecundado y	para	su desarrollo hasta la fase terminal. Es decir, a lo largo de la transformación		
cíclica, llega un momento, después de la ovulación, en que todo está como a punto	para	la continuación del embarazo. Por eso se ha dicho, gráficamente, que la primera		
eso se ha dicho, gráficamente, que la primera mitad del ciclo es la preparación	para	el embarazo, y la segunda parte es la situación adecuada para el comienzo del embarazo		
es la preparación para el embarazo, y la segunda parte es la situación adecuada	para	el comienzo del embarazo.		
determinados órganos y pasan al aparato circulatorio, disueltas en la sangre circulante,	para	actuar sobre los diversos tejidos, en los cuales producen sus efectos propios y		
Las hormonas ováricas, a medida que se forman van pasando a la sangre	para	actuar sobre los centros nerviosos del hipotálamo y, en consecuencia, sobre la		
conjunto —que representa una unidad funcional, cuya integridad es necesaria	para	la normalidad— constituye un complicado mecanismo, de función esencial y		
	Para	describir la sucesión de fenómenos del ciclo se suele comenzar a partir de la menstruación		

Abbildung 8: CREA – Anzeige der Konkordanzen für *para* als Präposition in der annotierten Version des CREA.²⁷⁵

In der annotierten Version des CREA ist es ebenso wie in der alten Version möglich, die Suche auf bestimmte geographische, zeitliche, thematische oder mediale Bereiche einzugrenzen, wobei im Hinblick auf das Medium hinzugefügt werden muss, dass bislang keine mündlichen Texte verfügbar sind (Stand: August 2018).²⁷⁶ Neu ist jedoch, dass mittels der Option *Proximidad* auch ganze Folgen von Lexemen gesucht werden können, indem der gewünschte Abstand zwischen den Lexemen angegeben und bei der Suche berücksichtigt werden kann. Zudem können Kollokationen anhand der in Kapitel 1.4.8. vorgestellten statistischen Verfahren MI, *LL-Simple* und *T-Score* ermittelt werden.

Um Unklarheiten zu vermeiden, wird die alte Version des CREA im Folgenden als CREA (Version 2015) bezeichnet. Sofern lediglich der Name CREA genutzt wird, bezieht er sich auf die neue Version. Da es sich allerdings grundsätzlich um annähernd dieselben Texte handelt, sind die Ergebnisse vergleichbar. Die chronologische Grenze der neuen, annotierten Version des CREA wurde jedoch auf das Jahr 2000 gelegt, da das CORPES, welches im folgenden Kapitel beschrieben wird, Texte ab dem Jahr 2001 enthält.²⁷⁷ Somit umfasst das CREA (Version 2015) zusätzlich Texte aus den Jahren 2001 bis 2004, welche im neuen CREA fehlen, da sie dem CORPES zugeschrieben wurden.

2.1.1.2. CORPES – *Corpus del Español del Siglo XXI*

Das *Corpus del Español del Siglo XXI*, kurz CORPES, ist ebenso wie das CREA ein Referenzkorpus. Es handelt sich hierbei um ein sehr aktuelles Korpus, das geschriebene und mündliche Texte aus Spanien, Lateinamerika, den Philippinen und Äquatorialguinea aus den Jahren 2001 bis 2012 enthält. Das Korpus befindet sich noch

²⁷⁵ Real Academia Española (ed.): *CREA. Versión anotada. Entrada Compleja*. Online Publikation: <http://web.frl.es/CREA/org/publico/pages/consulta/entradaCompleja.view> (17.04.2018).

²⁷⁶ Vgl. Real Academia Española (ed.): *CREA. Versión Anotada*. Op. cit. (30.08.2018).

²⁷⁷ Vgl. Real Academia Española (ed.): *CREA. Versión Anotada*. Op. cit. (03.12.2017).

im Aufbau, es ist jedoch seit Dezember 2013 eine kostenlose Testversion im Internet nutzbar (<http://web.frl.es/CORPES/view/inicioExterno.view>). Die derzeitige Version 0,91 (seit Dezember 2018) umfasst ungefähr 286 Mio. Formen.²⁷⁸ Die vorherige Version 0.9 (seit Juli 2018) umfasste knapp 277 Mio. Lexeme und deren Vorgänger, die Version 0.83 war von Juni 2016 bis September 2018 verfügbar und bestand aus 237.678 Dokumenten mit mehr als 225 Mio. Lexemen.²⁷⁹ Letztendlich sollte im Jahr 2018 das endgültige Korpus mit 400 Mio. Lexemen veröffentlicht werden, was jedoch nicht erreicht wurde.²⁸⁰ Es bleibt abzuwarten, wann das CORPES in seiner endgültigen Version veröffentlicht werden wird. In den Jahren 2016 und 2017 kamen ungefähr 50 Mio. Lexeme hinzu. Im Dezember 2018 zählte das CORPES 285.000 Dokumente mit ungefähr 286 Mio. Formen, es kamen also seit der Version im Juli 2018 zehn Mio. Formen dazu.²⁸¹

Die Erweiterung des CORPES wird dabei vom *Banco Santander* finanziert. Die *Real Academia Española* und der *Banco Santander* haben ihr Abkommen aus dem Jahr 2007 bis zum Jahr 2018 erweitert, wobei der Generaldirektor den finanziellen Beitrag von 700.000 Euro zur weiteren Entwicklung des CORPES zugesichert hat.²⁸²

Es handelt sich beim CORPES um ein sogenanntes halboffenes Korpus, das heißt, dass es nach einem vorgegebenen Plan erweitert wird (siehe Kapitel 2.7.). Dabei entfallen 90% der Texte auf die geschriebene und 10% auf die gesprochene Sprache. Die geschriebenen Texte werden zu jeweils 40% aus Büchern und Zeitungen entnommen, 7,5% aus dem Internet und 2,5% sind verschiedener Herkunft. Von allen Texten stammen 70% aus Lateinamerika und 30% aus Spanien, was eine deutliche geographische Verschiebung des Schwerpunktes auf Lateinamerika bedeutet. Hierbei wurde Lateinamerika in verschiedene linguistische Zonen eingeteilt, wobei hervorzuheben ist, dass das CORPES das erste akademische Korpus ist, welches auch Texte aus Äquatorialguinea und den Philippinen enthält. Die Texte sind thematisch in fiktionale und nicht-fiktionale Texte unterteilt. Im Korpus ist es also möglich, zum einen nach

²⁷⁸ Vgl. Real Academia Española (ed.): *Banco de datos. CORPES*. Online Publikation: <http://www.rae.es/recursos/banco-de-datos/corpes-xxi> (19.06.2019).

²⁷⁹ Vgl. Real Academia Española (ed.): *CORPES*. Online Publikation: <http://web.frl.es/CORPES/view/inicioExterno.view> (27.08.2018).

²⁸⁰ Vgl. Real Academia Española (ed.): *Banco de datos. CORPES*. Op. cit. (24.10.2016).

²⁸¹ Vgl. Real Academia Española (ed.): *Banco de datos. CORPES*. Op. cit. (19.06.2019).

²⁸² Vgl. Real Academia Española (ed.): *La RAE y Banco Santander renuevan su convenio para el CORPES XXI*. Online Publikation: <http://www.rae.es/noticias/la-rae-y-banco-santander-renuevan-su-convenio-para-el-corpes-xxi> (14.12.2016).

2. Kategorisierung der Korpora

Romanen, Erzählungen, Drehbüchern und Theaterstücken und zum anderen nach Texten aus Bereichen wie Kunst und Kultur, Wissenschaft und Technologie, Freizeit und Alltagsleben, Sozialwissenschaften und Glauben, Gesundheit sowie Politik, Wirtschaft und Justiz zu suchen.²⁸³ Zudem können die zu durchsuchenden Texte näher definiert werden, sie lassen also beispielsweise eine Suche nach Interviews, Magazinen, E-Mails, Nachrichten in sozialen Netzwerken oder Reportagen zu.²⁸⁴

Des Weiteren ist es im CORPES ebenso wie in der annotierten Version des CREA möglich, mittels der drei statistischen Verfahren MI, *LL-Simple* und *T-Score* Kollokationen zu ermitteln (siehe Kapitel 1.4.8.). Seit der Version 0.8 (April 2015) ist es sogar möglich, die Ausschnitte aus mündlichen Texten anzuhören, und es kann zudem nach grammatikalischen Kategorien gesucht werden, da das CORPES morphosyntaktisch annotiert und lemmatisiert ist.²⁸⁵ Seit der aktuellen Version 0.91 vom Dezember 2019 ermöglicht das Korpus dem Nutzer zudem, Audiodateien herunterzuladen oder sich Videos der Dokumente anzusehen.²⁸⁶ Die Benutzeroberfläche des CORPES ist so gestaltet wie diejenige der annotierten Version des CREA:

The screenshot shows the CORPES search interface. At the top is a navigation bar with links: **Concordancias**, **Coapariciones**, **Configuración**, **Ayuda**, **Modo de cita**, and **Sugerencias**. Below this is a search form with the following fields and options:

- Lema:** A text input field containing the word "ordenador".
- Forma:** A text input field.
- Clase de palabra:** A dropdown menu set to "sustantivo".
- ☐ **Grafía original**
- Número:** A dropdown menu.
- Género:** A dropdown menu.
- Tipo:** A dropdown menu.
- Subcorpus:** A dropdown menu set to "Limpiar".
- Título:** A text input field.
- Autor:** A text input field.
- Fecha de clasificación:** A date range selector.
- Origen:** A dropdown menu with options: (Todos), América, España, Filipinas.
- Medio:** A dropdown menu with options: (Todos), Escrito, Oral.
- Tipología:** A dropdown menu with options: (Todos), Debate, Discurso, Entrevista.
- Sexo:** A dropdown menu with options: (Todos), Hombre, Mujer, No indicado.
- Marcas:** A section with checkboxes for ☐ Cursiva, ☐ Negrita, ☐ Subrayado, and ☐ Versalitas.

At the bottom of the form are three buttons: **Concordancia**, **Estadística**, and **Nueva consulta**.

Abbildung 9: CORPES – Suche nach dem Substantiv *ordenador* im mündlichen Subkorpus.²⁸⁷

Der Nutzer kann auch hier aus allen verfügbaren Texten ein Subkorpus wählen, in welchem gesucht werden soll. Im vorliegenden Beispiel soll das Substantiv *ordenador* in allen mündlichen Texten gesucht werden. Bei dieser Suche werden die Konkordanzen

²⁸³ Vgl. Real Academia Española (ed.): *CORPES. Descripción del sistema de codificación*. Online Publikation: http://www.rae.es/sites/default/files/CORPES_Sistema_de_codificacion_12_2015.pdf (24.10.2016).

²⁸⁴ Vgl. Real Academia Española (ed.): *CORPES*. Op. cit. (17.04.2018).

²⁸⁵ Vgl. Real Academia Española (ed.): *CORPES*. Op. cit. (17.04.2018).

²⁸⁶ Vgl. Real Academia Española (ed.): *Banco de datos. CORPES*. Op. cit. (19.06.2019).

²⁸⁷ Real Academia Española (ed.): *CORPES*. Op. cit. (28.08.2018).

ebenso angezeigt wie im CREA, es sind jedoch zusätzlich am linken Rand der Konkordanzen Lautsprecher zu sehen, die anzeigen, dass für den jeweiligen Ausschnitt eine Audiodatei vorliegt. Durch einen Klick auf eines der blau markierten Ergebnisse wird der sprachliche Kontext dieses Beispiels im unteren Bereich der Benutzeroberfläche angezeigt:

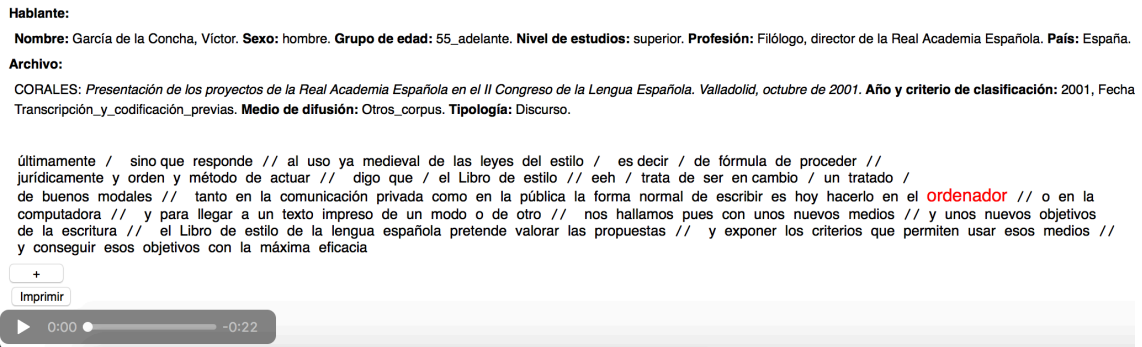


Abbildung 10: CORPES – Anzeige des Suchbegriffs *ordenador* im jeweiligen sprachlichen Kontext inklusive Audiodatei.²⁸⁸

Zusätzlich zum sprachlichen Kontext werden auch alle verfügbaren Informationen über den ausgewählten Ausschnitt aufgelistet, und die begleitende Audiodatei kann vom Nutzer abgespielt werden.

2.1.1.3. CdE:Orig – Der aktuelle Teil des *Corpus del Español (Género/Histórico)*

Das so bezeichnete *Corpus del Español* (CdE), welches von Mark Davies erstellt wurde, besteht seit dem Sommer 2016 aus zwei unterschiedlichen Korpora.²⁸⁹ Das ursprüngliche Korpus, welches nun *Corpus del Español: Género/Histórico* (kurz: CdE:Orig) heißt, wurde im Jahr 2002 online veröffentlicht (<https://www.corpusdelespanol.org/hist-gen/>). Es handelt sich dabei im Grunde um ein historisches Korpus (siehe auch Kapitel 2.2.1.3.), welches 100 Mio. Lexeme aus über 20.000 Texten vom 13. bis 20. Jahrhundert umfasst.²⁹⁰ Allerdings enthält es allein 20 Mio. Lexeme aus dem späten 20. Jahrhundert.²⁹¹ Dieser aktuelle Teil ist ausgeglichen bezüglich der Zusammenstellung der Texte aus unterschiedlichen Genres (mündliche,

²⁸⁸ Real Academia Española (ed.): *CORPES. Entrada Compleja*. Online Publikation: <http://web.frl.es/CORPES/org/publico/pages/consulta/entradaCompleja.view> (28.08.2018).

²⁸⁹ Vgl. Davies, Mark: *Corpus del Español. Corpora*. Online Publikation: <http://www.corpusdelespanol.org/x.asp> (21.11.2016).

²⁹⁰ Vgl. Davies, Mark: *Corpus del Español. Introducción*. Online Publikation: <http://www.corpusdelespanol.org/> (20.05.2011).

²⁹¹ Vgl. Davies/ Face (2006): op. cit., S. 134.

fiktionale, journalistische und akademische Texte) und kann somit als Referenzkorpus bezeichnet werden.²⁹² Das Subkorpus des aktuellen Spanisch ist zwar ein Teil des CdE:Orig, es kann jedoch separat genutzt werden, wenn die zu durchsuchende Zeitperiode auf das 20. Jahrhundert eingegrenzt wird. Somit kann es durchaus als eigenständiges Referenzkorpus genutzt werden. Wenn in der vorliegenden Arbeit der Name CdE:Orig verwendet wird, so ist immer nur dieser moderne Teil des CdE:Orig gemeint, aufgrund der Tatsache, dass von diesem aktuellen Subkorpus häufiger die Rede sein wird als vom gesamten historischen Korpus. Sofern speziell vom historischen Korpus gesprochen wird, soll es als CdE:Orig (Gesamt) bezeichnet werden.

Das CdE:Orig besteht zu zwei Dritteln aus geschriebenen Texten, ein Drittel mit immerhin 6.750.000 Lexemen repräsentiert das gesprochene Spanisch. Ungefähr die Hälfte des gesprochenen Korpus ist durch Transkriptionen natürlicher Sprache repräsentiert, wie die 2,3 Mio. Lexeme aus dem *Habla Culta* Korpus, das Gespräche mit Sprechern aus elf verschiedenen Ländern beinhaltet und die 1 Mio. Lexeme aus dem *Corpus Oral de Referencia*, welches unter anderem Transkriptionen von Gesprächen, Vorträgen, Reden und Sportberichterstattungen enthält. Das geschriebene Korpus besteht zur Hälfte aus literarischen und zur Hälfte aus nicht-literarischen Texten, wie Zeitungsartikeln, Aufsätzen, Enzyklopädien, Briefen und humanistischen Texten. Das Korpus ist auch in Bezug auf die Herkunft der Texte ausgeglichen, da ungefähr 43% auf Spanien entfallen, während 57% aus Lateinamerika stammen.²⁹³

Des Weiteren sind alle Texte aus der Zeit von 1970 bis 2000, wobei die meisten Texte aus den 1990er Jahren stammen.²⁹⁴ Im Jahr 2002 modifizierten Mark Davies und Douglas Biber das Korpus des aktuellen Spanisch, um sicherzustellen, dass alle Texte frühestens aus den 1970er Jahren sind und es eine gerechte Aufteilung zwischen gesprochenen, schriftlich-fiktiven und schriftlich-nicht-fiktiven Texten gibt (genauere Angaben zur Herkunft der Texte des CdE:Orig und der Einteilung der Texte des Korpus des aktuellen Spanisch nach Genres und Registern finden sich in den Anhängen 5 und 6).²⁹⁵ Im Jahr 2007 wurde das CdE:Orig modifiziert, und die Suchmaske sowie die

²⁹² Vgl. Davies: *Corpus del Español. Introducción*. Op. cit. (20.05.2011).

²⁹³ Vgl. Davies (2006): op. cit., S. 2.

²⁹⁴ Vgl. Davies (2006): op. cit., S. 2.

²⁹⁵ Vgl. Davies (2005): op. cit., S. 107.

Struktur wurden stark verändert.²⁹⁶ Im Jahr 2016 wurde das Interface aufgrund der Überarbeitung des CdE:Orig und der Gestaltung des zweiten Korpus, namens *Corpus del Español: Web/Dialectos* (kurz: CdE:Nuevo, siehe auch Kapitel 2.6.1.), erneuert.²⁹⁷

In den Korpora CdE:Orig und CdE:Nuevo sind zunächst 10 bis 15 Eingaben ohne vorherige Registrierung möglich. Danach muss der Benutzer sich anmelden und ein Profil erstellen. Die Verwendung des Korpus ist zwar kostenlos, die Anmeldung unter Erstellung eines Profils ist aber unumgänglich. Ein angemeldeter Nutzer, der keinen Beitrag zahlt, hat pro Tag 50 Suchanfragen zur Verfügung. Darüber hinaus gibt es die Möglichkeit, durch die Zahlung von 10 \$ pro Monat, 30 \$ für ein Jahr, 55 \$ für zwei oder 75 \$ für drei Jahre einen *premium account* zu beantragen.²⁹⁸

Im lemmatisierten und morphosyntaktisch annotierten CdE:Orig kann unter anderem nach Synonymen gesucht oder zwei Lexeme können miteinander verglichen werden, beispielsweise *empezar* und *comenzar*:

Abbildung 11: CdE:Orig – Suche nach Substantiven, die zusammen mit den Verben *empezar* oder *comenzar* erscheinen.²⁹⁹

Anhand der Ergebnisse kann bereits auf einen Blick erkannt werden, dass *empezar* vorwiegend in Kontexten genutzt wird, in denen ‚jemand etwas beginnt‘, während

²⁹⁶ Vgl. Davies, Mark: *Corpus del Español. Ayuda: Información técnica y créditos*. Online Publikation: <http://www.corpusdelespanol.org/> (28.05.2011).

²⁹⁷ Vgl. Davies, Mark: *Corpus del Español – Género/Histórico*. Online Publikation: <http://www.corpusdelespanol.org/hist-gen/> (19.12.2016).

²⁹⁸ Vgl. Davies, Mark: *Corpus.byu.edu. Premium account*. Online Publikation: <https://corpus.byu.edu/premium.asp?c=y> (04.03.2018).

²⁹⁹ Davies: *Corpus del Español – Género/Histórico*. Op. cit. (17.04.2018).

2. Kategorisierung der Korpora

comenzar vielmehr dann gebraucht wird, wenn ‚etwas von sich aus beginnt‘. Es handelt sich hierbei allerdings lediglich um Tendenzen:

PALABRA 1 (P1): EMPEZAR (1.91)						PALABRA 2 (P2): COMENZAR (0.52)					
	PALABRA	P1	P2	FREC1/FREC2	PUNT		PALABRA	P2	P1	FREC2/FREC1	SCORE
1	COSA	4	0	8.0	4.2	1	DÉCADA	5	0	10.0	19.1
2	TRABAJOS	7	1	7.0	3.7	2	EJERCICIOS	4	0	8.0	15.3
3	DISCURSO	5	1	5.0	2.6	3	SESIÓN	4	1	4.0	7.6
4	ENTREVISTA	4	1	4.0	2.1	4	CAMINO	4	1	4.0	7.6
5	ARTÍCULO	4	1	4.0	2.1	5	ACTO	4	1	4.0	7.6
6	TARDE	4	1	4.0	2.1	6	ENSAYOS	5	2	2.5	4.8
7	PROGRAMA	7	2	3.5	1.8	7	OBRA	6	3	2.0	3.8
8	CONVERSACIÓN	7	2	3.5	1.8	8	SIGLO	4	2	2.0	3.8

Abbildung 12: CdE:Orig – Ergebnis der Suche nach Substantiven, die zusammen mit *empezar* oder *comenzar* verwendet werden.³⁰⁰

Mittels der Konkordanzen, zu denen man durch einen Klick auf die Anzahl der Treffer des jeweiligen Lexems in den Spalten P1 beziehungsweise P2 gelangt, wird diese erste Vermutung bestätigt, da sich hier Phrasen wie *puedo empezar una cosa* oder *para empezar los trabajos* auf der einen Seite und *al comenzar la década* oder *antes de comenzar el acto segundo* auf der anderen Seite finden.

Seit der Überarbeitung der Benutzeroberfläche des CdE:Orig im Jahr 2016 sind zwei verschiedene Versionen des Korpus online zugänglich. Der Unterschied zwischen den beiden Versionen des CdE:Orig ist allerdings nicht so groß wie derjenige zwischen CREA und CREA (Version 2015), da es sich hierbei um keine Änderung in den Annotationen handelt, sondern lediglich um ein neues Interface. Die neu gestaltete Benutzeroberfläche bietet einige Vorteile im Vergleich zur vorherigen Version: Sie läuft auch auf mobilen Geräten und ist sogar auf Smartphones übersichtlich gestaltet. Sie ist, laut Aussage von Davies, einfacher zu bedienen, da nur noch diejenigen Wahlmöglichkeiten erscheinen, welche ausgewählt wurden und der Nutzer dadurch nicht von den zusätzlichen Einstellungen überfordert wird. Die Informationen zur Funktionsweise des Korpus wurden überarbeitet und benutzerfreundlicher gestaltet, die Suchsyntax wurde vereinfacht (wobei weiterhin auch die alte Suchsyntax verwendet werden kann) und die Möglichkeit der Erstellung virtueller Korpora für individuelle Suchen wurde eingeführt.³⁰¹

Derzeit existieren also zum einen zwei verschiedene Teile des CdE – das CdE:Orig und das CdE:Nuevo. Zum anderen gibt es für das CdE:Orig ein altes und ein neues

³⁰⁰ Davies: *Corpus del Español – Género/Histórico*. Op. cit. (17.04.2018).

³⁰¹ Vgl. Davies, Mark: *Corpus del Español. New interface*. Online Publikation: <http://www.corpusdelespanol.org/interface2016.asp> (16.01.2017).

Interface. Sofern bestimmte Suchoptionen für alle Korpora und Benutzeroberflächen gelten, wird in dieser Arbeit generell vom CdE gesprochen. Sollte sich eine Suche auf ein bestimmtes Korpus beziehen, wird dieses mit CdE:Orig oder CdE:Nuevo betitelt und die Nutzung der alten Benutzeroberfläche aus dem Jahr 2008 wird mit der Angabe CdE:Orig (Version 2008) gekennzeichnet, um im jeweiligen Fall klarzustellen, welches Korpus in welcher Version verwendet wird.

2.1.2. Spezialkorpora

Dem Referenzkorpus steht das Spezialkorpus gegenüber. Ein Spezialkorpus wird laut Teubert/Čermáková erstellt, wenn der Fokus nicht auf der jeweiligen Sprache in ihrer Gesamtheit liegt, sondern auf einem besonderen linguistischen Phänomen. Spezialkorpora sind oft relativ klein und enthalten weniger als eine Mio. Lexeme, sie können jedoch auch größer sein.³⁰²

Nach Teubert/Čermáková gibt es keine allgemeingültige Anleitung zur Erstellung eines Spezialkorpus. Es müssen nur einige Hypothesen aufgestellt werden, um das jeweilige Spezialkorpus zu definieren. Allerdings ist es möglich, dass die Hypothesen im Laufe der Untersuchungen modifiziert werden müssen. Daher kann es sein, dass das Korpus erweitert werden muss, was laut den Autoren kein Problem darstellt, da es das Produkt des Sprachwissenschaftlers ist und somit auch jederzeit von diesem verändert werden kann.³⁰³

Die historischen Korpora *Biblia Medieval* und ADMYTE (*Archivo Digital de Manuscritos y Textos Españoles*), welche sich ausschließlich mit dem Spanischen im Mittelalter befassen, Ersteres sogar lediglich mit Übersetzungen der Bibel, können als Spezialkorpora betrachtet werden. Sie werden in den Unterkapiteln 2.2.1.4. beziehungsweise 2.8.1.2. genauer beschrieben.

Zudem können auch Lernerkorpora als Spezialkorpora betrachtet werden, beispielsweise das *Corpus Oral de Español como Lengua Extranjera (ELE)*, kurz CORELE. Sie sollen schließlich nicht das in spanischsprachigen Ländern gesprochene Spanisch darstellen, sondern die von Spanischlernern gesprochene Sprache. Auf diese Korpora wird im Abschnitt ‚Funktionalität‘ detailliert eingegangen, da das Anliegen für

³⁰² Vgl. Teubert/Čermáková (2007): op. cit., S. 68.

³⁰³ Vgl. Teubert/Čermáková (2007): op. cit., S. 69.

2. Kategorisierung der Korpora

die Erstellung ein anderes ist als beispielsweise bei Referenzkorpora (siehe Kapitel 2.2.2.).

Des Weiteren sind mehrsprachige Parallelkorpora wie das CLUVI oder das *Corpus Trilingüe Paralelo Grial* (siehe Kapitel 2.3.) und einige Korpora, die sich ausschließlich mit gesprochener Sprache beschäftigen, wie das C-ORAL-ROM (siehe Kapitel 2.4.2.2.), als Spezialkorpora zu sehen, da in diesen Fällen der Fokus ebenso wenig auf der spanischen Sprache in ihrer Gesamtheit liegt. Auch die multimodalen Korpora *Sacodeyl* und *Backbone Spanish* (siehe Kapitel 2.4.3.) können als solche bezeichnet werden, weil sie Videosequenzen von Schülern zwischen 13 und 18 Jahren zu Alltagsthemen bereitstellen. Auf all diese Korpora wird jedoch in gesonderten Unterbereichen der Korpuskategorisierung, welche ihrer Beschaffenheit und Zielsetzung besser entsprechen, eingegangen werden.

Dennoch sollen hier zwei Spezialkorpora beschrieben werden. Zum einen handelt es sich dabei um das Korpus COLA, ein Korpus der spanischen Jugendsprache, und zum anderen um das *Spanish in Texas Corpus*, ein multimodales Korpus mit spanischen und bilingual spanisch-englischen Sprachausschnitten.

2.1.2.1. COLA – *Corpus Oral de Lenguaje Adolescente*

Das *Corpus Oral de Lenguaje Adolescente* (COLA) ist ein Korpus der Jugendsprache, das gesprochene Sprachdaten Jugendlicher zwischen 13 und 19 Jahren aus Madrid sowie aus den lateinamerikanischen Metropolen Buenos Aires und Santiago de Chile enthält. Das Korpus ist orthographisch transkribiert worden und erlaubt Untersuchungen zum kommunikativen Stil von Jugendlichen sowie zum Einfluss anderer Sprachen auf die Jugendsprache.³⁰⁴ Es handelt sich dabei um aufgenommene, spontane Sprachausschnitte aus verschiedenen Schulen und Jugendclubs in den drei genannten Städten im Umfang von 500.000 Lexemen.³⁰⁵

Das Korpus, welches von Annette Myre Jørgensen von der Universität Bergen erstellt wurde, ist kostenlos im Internet verfügbar (<http://clu.uni.no/humfak/cola/>), die

³⁰⁴ Vgl. Jørgensen, Annette Myre: *El corpus COLA*. Online Publikation: http://www.colam.org/om_prosj-espanol.html (24.02.2018).

³⁰⁵ Vgl. Meurer, Paul: *Corpuscle. COLA*. Online Publikation: <http://clarino.uib.no/korpuskel/clarino-metadata?session-id=244347694754873&corpus=cola&default-corpus=cola&resource=cola> (24.02.2018).

Suchmaske ist jedoch lediglich auf norwegisch zugänglich. Die Texte können unter anderem nach den Informanten und deren Alter sortiert werden, und es kann nach bis zu drei Lexemen gesucht werden.³⁰⁶

Auf der Startseite der Daten aus Madrid befinden sich Links zum Textarchiv, zur Suchmaske, zu den Daten aus Buenos Aires und Santiago de Chile sowie eine alphabetische Wortliste und eine Frequenzliste.³⁰⁷ Im Textarchiv sind die Texte aller Informanten inklusive der Tonspur verfügbar, wobei Letztere aus einzelnen Sequenzen zur jeweiligen Textzeile besteht.

T L 10 ASPEAKERS: <Un chico y una chica conversan con otro chico/>
T L 10 MABPE2G04: es u u una sensación de decir hostias es que me caigo y me mato sabes/
T L 10 pero es que lo estoy saltando sabes/
T L 10 luego sí cuando caes al otro lado
T L 10 MABPE2G04: 1[te pegas una hostia]
MABPE2G02: 1[te matas je]
T L 10 MABPE2G04: 2[de la vida sabes/]
MABPE2G02: 2[je je je je]
T L 10 MABPE2G04: pero dices es que me salto una puta carretera sabes/
T L 10 es la hostia..y todo por la pista máas divertida
T L 10 MABPE2G02: es queee
T L 10 MABPE2G04: esquiar mola
T L 10 esquiaaar.. qué te gusta más/ esquiar en cama o esquiar en nieve/
T L 10 a que no sabrías decírmelo/

Abbildung 13: COLA – Ausschnitt aus dem Transkript eines Gesprächs zwischen drei Jugendlichen.³⁰⁸

Im Textarchiv können die Audiodateien allerdings nur abschnittsweise abgespielt werden. Möchte sich der Nutzer einen vollständigen Text anhören, so muss er auf der Seite des Korpus dem Link *Conversaciones íntegras* folgen, um zum Archiv der Audiodateien zu gelangen.³⁰⁹

³⁰⁶ Vgl. Jørgensen, Annette Myre: *El corpus COLA. Suchmaske*. Online Publikation: <http://clu.uni.no/humfak/cola/s9.html> (24.02.2018).

³⁰⁷ Vgl. Jørgensen, Annette Myre: *El corpus COLA. Startseite*. Online Publikation: <http://clu.uni.no/humfak/cola/> (24.02.2018).

³⁰⁸ Jørgensen, Annette Myre: *El corpus COLA. MABPE2-06*. Online Publikation: <http://clu.uni.no/humfak/cola/txt-lyd/mabpe2-06.htm> (17.04.2018).

³⁰⁹ Vgl. Jørgensen, Annette Myre: *El corpus COLA. Corpus*. Online Publikation: <http://www.colam.org/transkripsjoner-espanol.html> (28.02.2018).

2.1.2.2. *Spanish in Texas Corpus*

Das *Spanish in Texas Corpus* ist ein multimodales Korpus mit spanischen und bilingual spanisch-englischen Sprachausschnitten aus Interviews und Gesprächen zwischen Sprechern aus unterschiedlichen Regionen in Texas. Es besteht derzeit aus über 500.000 Lexemen von 97 bilingualen, in Texas lebenden Sprechern. Auf der Homepage können Video- und Audiodateien, vollständige Transkripte sowie POS-Annotationen heruntergeladen werden.³¹⁰

Laut den Projektverantwortlichen ist Spanisch in Texas keine Fremdsprache, da es regelmäßig zu Hause, in der Arbeit sowie mit Freunden und Nachbarn gesprochen wird. Im Unterschied zu anderen Ländern, in denen die spanische Sprache verwendet wird, beherrschen die meisten spanischsprechenden Texaner auch Englisch. Daher soll das Korpus die spanische Sprache, wie sie heute in Texas gesprochen wird, repräsentieren, und Studenten, Lehrern sowie generell allen Interessierten die Möglichkeit bieten, diese spanische Sprachvariation zu erkunden. Ein Ziel des Projekts ist es deshalb, der Allgemeinheit lokale Varietäten des Spanischen sowie deren Sprecher als Quelle für Informationen über Sprache und Kultur zu zeigen.³¹¹

Zu jedem Interview werden auf der Übersichtsseite das Geschlecht, das Alter, der Geburtsort, der Ort und das Jahr des Interviews, die Dauer des Videos und die Anzahl der Lexeme angegeben. Anhand dieser Informationen kann im Folgenden ein Video ausgewählt werden, wobei es sich um relativ lange Dateien zwischen neun Minuten und über einer Stunde handelt. Auf der Seite des jeweiligen Interviews finden sich neben den herunterladbaren Dateien, dem Video und einem Transkript zusätzliche Informationen über den Sprachgebrauch und die Spracherfahrung:

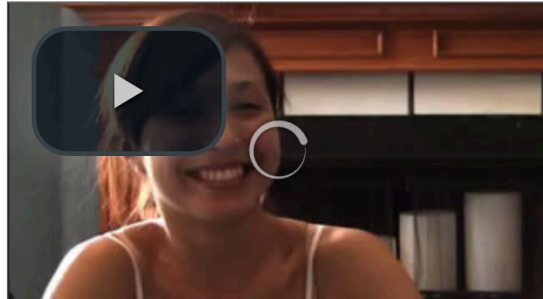
³¹⁰ Vgl. Bullock, Barbara E./ Toribio, Almeida Jacqueline (2013): *Spanish in Texas Corpus. About the Corpus*. Online Publikation: <http://corpus.spanishintexas.org/en/about-corpus> (06.10.2017).

³¹¹ Vgl. Bullock, Barbara E./ Toribio, Almeida Jacqueline (2013): *Spanish in Texas. About the Project*. Online Publikation: <http://spanishintexas.org/about-the-project/> (06.10.2017).

AF004_1991_EP_SU2011_AD

Interview

WORD COUNT: 4832 DURATION: 0:35:47



VIDEO	AUDIO	CAPTIONS	TAGS
mp4	wav	vtt srt	csv

Profile Language Use Language Experience

ELEMENTARY SCHOOL: Spanish and English
MIDDLE SCHOOL: English
HIGH SCHOOL: English
IN CHILDHOOD WITH PARENTS: Mostly Spanish
IN CHILDHOOD WITH SIBLINGS: Spanish and English
NOW WITH PARENTS: Spanish and English
NOW WITH SIBLINGS: Mostly English
NOW WITH FRIENDS: Mostly English

TRANSCRIPT

>>i: Muy buenas tardes. Estamos aquí con Ceci. Ceci muchas gracias por hacer esta entrevista.

>>s: De nada.

>>i: Ceci, ¿cuándo y dónde naciste?

>>s: Nací en agosto veintisiete... uhm... mil novecientos noventa y uno, en El Paso, Texas.

>>i: ¿Y dónde creciste y cómo era donde creciste?

Abbildung 14: *Spanish in Texas Corpus* – Interview mit einer 20-jährigen Frau aus El Paso.³¹²

Das Korpus ist im Internet kostenlos zugänglich, wobei der Nutzer sich zuvor registrieren und die Nutzungsbedingungen akzeptieren muss (<http://corpus.spanishintexas.org>). Zudem können im *SpinTX Video Archive* kostenlos und ohne vorherige Anmeldung ausgewählte Videos des *Spanish in Texas Corpus* angesehen werden (<http://www.coerll.utexas.edu/spintx/>). Dort finden sich hunderte kurzer Interviewausschnitte von ungefähr zwei Minuten Länge, welche nach Themen, Sprechern oder sprachlichen Besonderheiten in den Bereichen Grammatik, Pragmatik und Funktionen (beispielsweise ‚Personen beschreiben‘ oder ‚um Hilfe bitten‘) ausgewählt werden können. Auf der Seite des jeweiligen Interviewausschnitts finden sich hier allerdings nicht nur das Video mit Untertiteln, Angaben zu Sprechern und Themen sowie das Transkript, sondern auch ein interaktives Feld, in welchem man sich verschiedene sprachliche Einheiten anzeigen lassen kann.

³¹² Bullock, Barbara E./ Toribio, Almeida Jacqueline (2013): *Spanish in Texas Corpus. Interview AF004_1991_EP_SU2011_AD*. Online Publikation: <http://corpus.spanishintexas.org/en/interviews/af0041991epsu2011ad> (06.10.2017).

2. Kategorisierung der Korpora

The screenshot shows the SpinTX Video Archive interface. On the left, a transcript of an interview is displayed with several words highlighted in yellow: 'fue', 'es', 'era', 'eran', and 'es'. On the right, there is a sidebar with three tabs: 'Grammar', 'Vocabulary', and 'Other'. The 'Vocabulary' tab is selected, showing a list of verbs: 'ser', 'poner', 'haber', and 'tener'. Each verb has a checkbox next to it, and the 'ser' checkbox is checked. Above the list, there is a 'VIEW HIDE' button.

Abbildung 15: *SpinTX Video Archive* – Transkript eines Interviewausschnitts aus dem *Spanish in Texas Corpus*.³¹³

So kann sich der Nutzer beispielsweise alle Formen des Verbs *ser* im Transkript durch eine Markierung anzeigen lassen. Neben Einheiten zum Vokabular stehen hier auch Grammatikphänomene und Diskursmarker zur Auswahl.

2.1.3. Opportunistische Korpora

Da das Konzept der Repräsentativität von Korpora, wie es ein Referenzkorpus gemäß der gängigen Definition verlangt, relativ umstritten ist, führen Teubert/Čermáková das opportunistische Korpus als guten Mittelweg an. Diese Art von Korpora behauptet für sich nicht den Anspruch der Repräsentativität für eine Sprache und beruht auf der Annahme, dass jedes Korpus unausgeglichen ist. Bei der Arbeit mit opportunistischen Korpora wird strikt zwischen der Korpuserstellung und der Korpusanwendung getrennt. Bei der Erstellung wird versucht, so viele Korpora wie möglich zu vereinen. Die meisten Korpora sind Spezialkorpora, aber es können immer wieder auch Referenzkorpora darunter sein. Je größer das Korpus ist, desto besser, und je breiter die darin enthaltenen Spezialkorpora, desto mehr Linguisten mit unterschiedlichsten Vorhaben können mit dem opportunistischen Korpus arbeiten. Diese Korpora sind nicht in ihrer Größe beschränkt, und wann auch immer ein neues, interessantes Spezialkorpus gefunden wird, sollte es hinzugefügt werden.

Allerdings ist das beste opportunistische Korpus auch das am verständlichsten dokumentierte. Zunächst müssen die Genres und Textsorten bestimmt werden unter Berücksichtigung der folgenden zwei Aspekte: Zum einen muss beachtet werden, wonach mögliche zukünftige Benutzer suchen wollen. Es muss danach gefragt werden, welche Genres und Textsorten von Linguisten diskutiert und analysiert wurden. Zum anderen muss man sich fragen, welche Informationen man in den Korpora finden kann, die in das

³¹³ Bullock, Barbara E./ Toribio, Almeida Jacqueline (2013): *The Spanish in Texas Corpus Project*. COERLL. *SpinTX Video Archive*. Video 572. Online Publikation: <http://www.coerll.utexas.edu/spintx/video/572> (06.10.2017).

opportunistische Korpus integriert werden sollen. Man muss nachprüfen, ob sie Angaben über die genannten Kategorien Genre und Textsorte enthalten, die automatisch entnommen und in die Dokumentation aufgenommen werden können. Diesem Punkt sollte besondere Beachtung geschenkt werden, da die Integration eines neuen Korpus in ein bereits existierendes opportunistisches Korpus so automatisch wie möglich ablaufen sollte. Das ideale opportunistische Korpus enthält außerdem für jeden Text Informationen über dessen Erscheinungsdatum, den Namen des Autors, den Titel und alle Daten, die in der Bibliographie erscheinen sollten. Wenn Sprachwissenschaftler in einem ausreichend großen opportunistischen Korpus ihre Suchen durchführen, können sie somit zunächst in der Dokumentation nach geeigneten Texten suchen, mit dieser Auswahl arbeiten und die restlichen Texte vernachlässigen.³¹⁴

Lemnitzer/Zinsmeister geben allerdings eine noch freiere Definition dieser Korpora und führen an, dass opportunistische Korpora ohne jegliche Designkriterien zusammengestellt werden und alle verfügbaren Texte in sich vereinen. Diese Art der Korpora sei vor allem dort angebracht, wo allein die Datenmenge von Bedeutung sei, also besonders bei quantitativen Untersuchungen.³¹⁵ Generell spricht für die Verwendung eines solchen Korpus die Tatsache, dass kein Korpus wirklich repräsentativ ist (siehe auch Kapitel 1.1.2.).³¹⁶

Im Folgenden werden die zwei Korpora *Spa_news_2011_3M* und *Spa_newscrawl_2011* als Vertreter für opportunistische Korpora vorgestellt. Auch das *Corpus del Español Actual* (CEA) und das *Corpus periodístico Hemero*, welche als Korpora der geschriebenen Sprache in Kapitel 2.4.1. behandelt werden, sowie die Korpora CdE:Nuevo, *Google Books: Spanish* und *NOW Corpus: Spanish*, welche im Bereich ‚Größe‘ in Kapitel 2.6. beschrieben werden, gehören dieser Gattung an. Bei den letzten drei Korpora legt besonders die Definition von Lemnitzer/Zinsmeister die Einordnung in die Kategorie der opportunistischen Korpora nahe, da diese zwar nicht mehrere Korpora in sich vereinen (wie bei Teubert/Čermáková angegeben), bei ihrer Zusammenstellung jedoch hauptsächlich auf das Kriterium der Größe Wert gelegt wurde.

³¹⁴ Vgl. Teubert/Čermáková (2007): op. cit., S. 70-71.

³¹⁵ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 189.

³¹⁶ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 106-107.

2.1.3.1. *Spa_news_2011_3M (FCS Aggregator)*

Der so bezeichnete *FCS Aggregator* ist Teil der *CLARIN Federated Content Search* (CLARIN-FCS) Infrastruktur. Es handelt sich dabei um eine Benutzeroberfläche, die Suchen in den CLARIN-FCS Ressourcen ermöglicht. Dabei kann nach Korpora in verschiedenen Sprachen gesucht werden. Für die spanische Sprache hält der *FCS Aggregator* das Korpus *Spa_news_2011_3M* bereit. Es handelt sich dabei um ein spanisches Zeitungskorpus mit 3 Mio. Sätzen aus spanischen Zeitungstexten aus dem Jahr 2011.³¹⁷ Das Korpus ist frei im Internet verfügbar (<https://www.clarin.eu/contentsearch>).

Spa_news_2011_3M ist eine Korpussammlung aus 40 Subkorpora, welche Suchen in diesen Texten ermöglicht, keine Suchen mit *Wildcards* zulässt und lemmatisiert, aber nicht annotiert ist. Die Ergebnisse werden zunächst als Textausschnitte dargestellt, die Option *Key Word in Context* ist jedoch auch verfügbar. Die Suchergebnisse können in verschiedenen Formaten heruntergeladen und auf dem eigenen Computer gespeichert werden.³¹⁸

La ciudad ha sido sede de grandes manifestaciones en contra del gobierno del **presidente** Bashar Assad.

Andrei Rudakov, ex jugador del Spartak de Moscú y del Friburgo, fue nombrado **presidente** y es quien gestiona el día a día de la entidad.

Jordi Sevilla ha apostado por reformar el Senado "de una vez" y potenciar tanto la Conferencia de Presidentes, que reúne al **presidente** del Gobierno de España con los titulares de las comunidades autónomas, como otros instrumentos de cooperación.

"Urgimos al Capitolio a seguir al **presidente** Obama en su decisión de impedir que el dinero de los contribuyentes se destine a ayudar a la Argentina, un país que tiene recursos para pagar lo que debe y no lo hace", añade el anuncio.

El acuerdo reportará unos 165 millones de euros en cinco años a la tesorería azulgrana y el **presidente** , Sandro Rosell, ha asegurado en unas declaraciones a 'La Vanguardia' que dar marcha atrás al convenio supondría "replantear el tamaño de este club".

El jefe del Parlamento tunecino, Fouad Mebazza, asumió el sábado como **presidente** interino y de inmediato encomendó al primer ministro, Mohamed Ghannouchi, la tarea de formar un Gobierno de coalición.

Abbildung 16: *Spa_news_2011_3M* – Präsentation der Ergebnisse zum Suchbegriff *presidente*.³¹⁹

Leider ist das Korpus nicht annotiert, und der Nutzer kann sich nur einzelne Sätze anzeigen lassen. Zudem besteht es nur aus Zeitungstexten aus dem Jahr 2011 und ist daher auch in seinem Inhalt relativ beschränkt.

2.1.3.2. *Spa_newscrawl_2011 (Wortschatz Universität Leipzig)*

Ebenfalls anhand von Zeitungstexten aus dem Jahr 2011 wurde das *Spa_newscrawl_2011* erstellt, welches innerhalb der Korpussammlung *Wortschatz*

³¹⁷ Vgl. Weblicht (ed.2016): *FCS Aggregator*. Online Publikation: http://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/FCS_Aggregator (17.01.2017).

³¹⁸ Vgl. CLARIN (ed.): *FCS Aggregator*. Online Publikation: <https://www.clarin.eu/contentsearch> (17.01.2017).

³¹⁹ CLARIN (ed.): *FCS Aggregator. Suche nach ,presidente'*. Online Publikation: <https://spraakbanken.gu.se/ws/fcs/2.0/aggregator/#> (09.03.2017).

Universität Leipzig zugänglich und kostenlos nutzbar ist (http://corpora.uni-leipzig.de/de?corpusId=spa_newscrawl_2011). Es umfasst knapp 17 Mio. Sätze mit fast 2 Mio. *types*, verteilt auf über 391 Mio. *tokens*.³²⁰

Wort: **garrafal** Anzahl: 216 Rang: 59,835 Häufigkeitsklasse: 17

Siehe auch: [Garrafal](#)

▼ **Beispiele:** Ser un adulto con "actitudes predominantemente adolescentes"...

▲ **Kookkurrenzen:**

error (2,053), un (226), cometió (189), falla (102), arquero (50), pifia (49), es (43), balón (41), portero (41), alteren (38), Error (35), Furioso (32), fallo (29), gol (28), cometido (27), tiro (27), tras (27), Criminalística (27), Rápido (26), Huidobro (26), defensa (25), minutos (25), quien (24), planillas (24), comete (23), Sería (22), Usted (20), empate (20), aprovechó (19), tan (19), anotó (19), Rubio (17), disculpas (17), pelota (17), estratégico (16), libre (16), consiguió (16), remarcó (15), costó (15), Agregó (15), cumpliendo (15), debió (15), al (15), salida (14), área (13), Guillermo (13), capitán (13), triunfo (13), derecho (12), el (12), aprovechar (12), vencer (12), haber (12), lectores (12), político (11), salir (11), campaña (11), que (11), e (10), definitiva (10)

▼ **Signifikante linke Nachbarn:** error | falla | Error | tan | fallo

▼ **Signifikante rechte Nachbarn:** error | falla | . | " | al

Abbildung 17: *Spa_newscrawl_2011* – Ergebnis der Suche nach dem Adjektiv *garrafal*.³²¹

Als Ergebnis listet das Korpus neben der gesamten Anzahl, dem Rang und einer Angabe zur Häufigkeitsklasse nicht nur Beispielsätze auf, sondern auch Kookkurrenzen des Suchbegriffs mit der jeweiligen Trefferanzahl, signifikante linke und rechte Nachbarn sowie einen Graphen mit häufigen Kombinationen (siehe Kapitel 4.1.1. zur Diskussion des Begriffs ‚Kookkurrenz‘ im Rahmen der Abgrenzung vom Begriff ‚Kollokation‘).

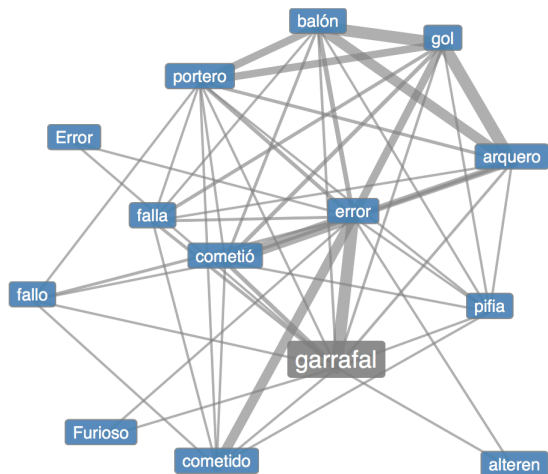


Abbildung 18: *Spa_newscrawl_2011* – Graph zum Adjektiv *garrafal*.³²²

³²⁰ Vgl. Universität Leipzig - Institut für Informatik - Abteilung ASV (1998-2017): *Wortschatz Universität Leipzig. Spa_newscrawl_2011*. Online Publikation: http://corpora.uni-leipzig.de/de?corpusId=spa_newscrawl_2011 (21.10.2017).

³²¹ Universität Leipzig - Institut für Informatik - Abteilung ASV (1998-2017): *Wortschatz Universität Leipzig. Spa_newscrawl_2011. Suche nach ‚garrafal‘*. Online Publikation: http://corpora.uni-leipzig.de/de/res?word=garrafal&corpusId=spa_newscrawl_2011 (21.10.2017).

³²² Universität Leipzig - Institut für Informatik - Abteilung ASV (1998-2017): *Wortschatz Universität Leipzig. Spa_newscrawl_2011. Suche nach ‚garrafal‘*. Op. cit. (21.10.2017).

2. Kategorisierung der Korpora

Bei dieser Darstellung werden bedeutende Kookkurrenzen angezeigt, wobei die Breite der Verbindungslinie den Grad der Relevanz andeutet. Dabei wird auch die Verbindung der anderen Nachbarn untereinander angezeigt. Irrelevante Kombinationen beispielsweise mit Artikeln oder Kopulaverben wie *ser* werden nicht aufgeführt.

Auf der Internetseite des Korpus *Spa_news_2011_3M* wird angegeben, dass die Quelle dieses Korpus die *Automatische Sprachverarbeitung Leipzig (ASV Leipzig)* ist.³²³ Aus diesem Grund ist anzunehmen, dass die Daten des *Spa_news_2011_3M* sich mit denjenigen des *Spa_newscrawl_2011* überschneiden beziehungsweise einen Ausschnitt daraus darstellen.

2.2. Funktionalität

Die Funktionalität, also der Zweck eines Korpus, liegt den anderen Kriterien zugrunde. Sie ergibt sich unter anderem aus den Designkriterien, der Korpusgröße, dem Umfang der Annotation und weiteren Entscheidungen im Hinblick auf die Korpuserstellung. Die ursprüngliche Funktionalität eines Korpus erklärt bestimmte Eigenschaften desselben, da sie die intendierte Nutzung widerspiegelt. Dennoch kann das Korpus später zu anderen Zwecken verwendet werden.³²⁴

Im Normalfall wird ein Korpus als „empirische Basis für linguistische und/oder computerlinguistische Forschung“ erstellt, was beispielsweise bei großen, opportunistischen Korpora der Fall ist.³²⁵ Diese Angabe bestimmt jedoch nicht direkt die Funktionalität desselben. Annähernd ebenso verhält es sich mit der Angabe „Sprachdokumentation“, da anhand der Korpora stets Sprache dokumentiert wird. Hiermit ist jedoch nicht gemeint, dass eine beliebige Sprache an sich dokumentiert wird, sondern dass das Korpus einen Sprachstand dokumentieren soll, der nicht mehr durch neue Datenerhebungen ersetzt werden kann, wie es unter anderem bei historischen Korpora der Fall ist.³²⁶

2.2.1. Historische Korpora

Einige Korpora beschäftigen sich demnach nicht mit dem Spanischen der Gegenwart, sondern beinhalten spanische Texte aus verschiedenen Jahrhunderten der

³²³ Vgl. CLARIN (ed.): *FCS Aggregator*. Op. cit. (17.01.2017).

³²⁴ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 102-103.

³²⁵ Lemnitzer/ Zinsmeister (2010): op. cit., S. 108.

³²⁶ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 108.

Vergangenheit. In diesen Bereich fallen zum einen das CORDE und das CDH der *Real Academia Española* sowie das CdE:Orig (Gesamt) von Mark Davies. Diese Korpora stechen dadurch hervor, dass sie Texte der gesamten Geschichte des Spanischen beinhalten, also von den Ursprüngen der spanischen Sprache bis ins 20. Jahrhundert, und sie stellen damit, nicht zuletzt wegen der sorgfältigen Auswahl der Texte nach Genres und Epochen, eine Art Referenzkorpus für das historische Spanisch dar. Zum anderen existieren kleinere Spezialkorpora, die vor allem zu Forschungszwecken im Bereich der Sprachgeschichte und der Literaturwissenschaft verwendet werden.

2.2.1.1. CORDE – *Corpus Diacrónico del Español*

Das *Corpus Diacrónico del Español* (CORDE) ist ein von der *Real Academia Española* im Jahr 2001 veröffentlichtes historisches Korpus.³²⁷ Es enthält Texte von den Ursprüngen des Spanischen bis zum jeweiligen Jahr, in dem das CREA beginnt, welches nur Texte der letzten 25 Jahre umfasst (siehe Kapitel 2.1.1.1.). Derzeit reicht das CORDE bis ins Jahr 1974, da das CREA Texte von 1975 bis 1999 (beziehungsweise bis 2004 in der Version von 2015) umfasst.³²⁸

Momentan zählt das CORDE 250 Mio. Lexeme.³²⁹ Es handelt sich dabei um ein Korpus der geschriebenen Sprache mit vollständigen Texten, die jedoch nicht angesehen oder heruntergeladen werden können.³³⁰ Ebenso wie das CREA ist es morphosyntaktisch annotiert.³³¹ Das CORDE ist im Internet kostenlos und frei zugänglich (<http://corpus.rae.es/cordenet.html>). In Bezug auf die geographische, zeitliche und gattungsspezifische Verteilung der Texte kann das CORDE als repräsentatives Korpus bezeichnet werden, wie im Folgenden zu sehen sein wird.³³²

Im Hinblick auf die geographische Verteilung beschäftigt sich das CORDE mit dem Spanischen in allen Teilen der Welt, in denen man Spanisch spricht oder in der

³²⁷ Vgl. Davies/ Face (2006): op. cit., S. 133.

³²⁸ Vgl. Real Academia Española (ed.): *CREA. Ayuda*. Op. cit. (14.05.2011).

³²⁹ Vgl. Real Academia Española (ed.): *Banco de datos. CORDE*. Online Publikation: <http://www.rae.es/recursos/banco-de-datos/corde> (23.01.2017).

³³⁰ Vgl. Real Academia Española (ed.): *CREA. Ayuda*. Op. cit. (14.05.2011).

³³¹ Vgl. Sánchez, Fernando/ Porta, Jordi/ Sancho, José Luis/ Nieto, Amalio/ Ballester, Almudena/ Fernández, Adelaida/ Gómez, Javier/ Gómez, Laura/ Raigal, Encarnación/ Ruiz, Rafael: *La anotación de los corpus CREA y CORDE*. Online Publikation: <http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/3485/1973> (21.12.2016).

³³² Vgl. Real Academia Española (ed.): *Banco de datos. CORDE*. Online Publikation: <http://www.rae.es/recursos/banco-de-datos/corde> (12.01.2017).

2. Kategorisierung der Korpora

Vergangenheit gesprochen hat. Aufgrund der diachronischen Perspektive des Korpus kommen dem peninsularen Spanisch 74% der Texte zu.³³³ Das lateinamerikanische Spanisch nimmt 25% davon ein und das sephardische Spanisch sowie weitere kleine Sprachgruppen zusammen 1% der Texte.³³⁴

Chronologisch gesehen gliedert sich das CORDE in drei große Etappen, die wiederum in mehrere Perioden unterteilt sind und bezüglich der Textverteilung über ungleiche prozentuale Anteile verfügen: Das Mittelalter mit 21% (Ursprünge des Spanischen bis 1250, 1250-1492), die *Siglos de Oro* mit 28 % (1493-1598, 1599-1713) und die zeitgenössische Epoche mit 51% (1714-1812, 1813-1898, 1899-1936 und 1937-1974).³³⁵

Bezüglich der Genres lässt sich das Korpus in zwei große Gruppen einteilen: Fiktion (44%) und Nicht-Fiktion (56%). Im Bereich Fiktion handelt es sich um Texte in Gedichtform und Prosa, die in die Subgattungen Lyrik, Epik und Drama unterteilt sind. Der nicht-fiktive Bereich ist eingeteilt in die Themen Didaktik (9%), Wissenschaft und Technik (15%), Religion (6%), Geschichte (13%), Rechtswesen (5%) sowie Gesellschaft, Presse und Öffentlichkeit (8%).³³⁶ Insgesamt stammen 97% der Texte aus Büchern und 3% aus der Presse.³³⁷ Genauere Informationen zur Aufteilung der Einträge nach Gattung und Epoche finden sich in Anhang 7.

Obwohl sich die beiden Korpora CORDE und CREA komplementieren und die Texte des zweiten Korpus im Laufe der Jahre in das erste übergehen, gibt es auch Unterschiede zwischen den Korpora aufgrund der besonderen Zusammensetzung des CORDE und des Charakters der Texte dieses Korpus (beispielsweise Gedichte, annotierte Texte, vorläufige Textfassungen und andere Kompositionen, die nicht Teil eines Werks selbst sind). Daher hat das CORDE ein Kennzeichnungssystem, das es ihm erlaubt, mittels entsprechender Markierungen unter den Texten die Gedichte, die Beiträge in

³³³ Vgl. Real Academia Española (ed.): *CREA. Ayuda*. Op. cit. (14.05.2011).

³³⁴ Vgl. Real Academia Española (ed.): *Banco de datos. CORDE. Tamaño y criterios de selección*. Online Publikation: [http://www.rae.es/rae/gestores/gespub000019.nsf/\(voAnexos\)/arch475E744872738671C125716500381CF8/\\$FILE/TamanyocriteriosCORDE.htm](http://www.rae.es/rae/gestores/gespub000019.nsf/(voAnexos)/arch475E744872738671C125716500381CF8/$FILE/TamanyocriteriosCORDE.htm) (19.05.2011).

³³⁵ Vgl. Real Academia Española (ed.): *CREA. Ayuda*. Op. cit. (14.05.2011).

³³⁶ Vgl. Real Academia Española (ed.): *CREA. Ayuda*. Op. cit. (14.05.2011).

³³⁷ Vgl. Real Academia Española (ed.): *Banco de datos. CORDE. Tamaño y criterios de selección*. Op. cit. (19.05.2011).

anderen Sprachen und die Notizen zum Text zu identifizieren.³³⁸ Textuelle Probleme, wie frühere Textfassungen, Schätzungen, Zensuren und schriftstellerische Freiheiten, sowie jegliches Eingreifen fremder Autoren wurden mit verschiedenen Etiketten versehen, die die Unterscheidung der Worte des ursprünglichen Autors von denen der intervenierenden Autoren ermöglichen.³³⁹

Das CORDE dient als Untersuchungsbasis sowohl den Forschern, die an der Existenz eines Lexems beziehungsweise eines Ausdrucks interessiert sind oder eine grammatikalische Studie durchführen wollen, als auch den Lexikographen zur Erstellung des *Nuevo diccionario histórico del español*.³⁴⁰ Die *Real Academia Española* nutzt das CORDE, um Lexeme zu dokumentieren, sie als antiquiert oder ungebräuchlich zu qualifizieren sowie um ihren Ursprung, ihren tatsächlichen Gebrauch und ihr erstes Erscheinen zu erörtern.³⁴¹

Bei der Suchmaske des CORDE, die im Übrigen derjenigen des CREA (Version 2015) entspricht, hat der Nutzer folgende Möglichkeiten: Zunächst kann er im Feld *Medio* ein bestimmtes Medium, also Bücher, Zeitungen, Zeitschriften, Gemischtes oder mündliche Texte, auswählen. Er kann einen Autor bestimmen, indem er einfach den Nachnamen eintippt, beispielsweise *Lope* für Lope de Vega. Im Feld *Obra* kann, mithilfe des ersten Lexems des Titels oder auch mehrerer Lexeme in der richtigen Reihenfolge, nach einem bestimmten Werk gesucht werden. Des Weiteren kann der Benutzer sowohl nach Lexemen suchen, die nur in einem bestimmten Jahr vorkommen, als auch in einer bestimmten Zeitspanne. Dazu verwendet er nur das erste Feld neben *Cronológico* für ein bestimmtes Jahr und beide Felder, um eine Periode, die mehrere Jahre umfasst, anzugeben. Außerdem kann im Feld *Geográfico* aus verschiedenen spanischsprachigen Ländern und im Feld *Tema* aus vorgegebenen Themen ausgewählt werden. Sofern in einem Feld nichts eingetragen oder keine Option ausgewählt wurde, werden immer alle Ergebnisse aus dem gesamten Korpus angezeigt.

³³⁸ Vgl. Real Academia Española (ed.): *CREA. Ayuda*. Op. cit. (14.05.2011).

³³⁹ Vgl. Real Academia Española (ed.): *Banco de datos. CORDE. Información*. Online Publikation: <http://www.rae.es/rae/gestores/gespub000019.nsf/voTodosporId/B4E26FC2520104D8C125716400455C06?OpenDocument&i=1> (19.05.2011).

³⁴⁰ Vgl. Real Academia Española (ed.): *CREA. Ayuda*. Op. cit. (14.05.2011).

³⁴¹ Vgl. Real Academia Española (ed.): *Banco de datos. CORDE*. Op. cit. (12.01.2017).

2. Kategorisierung der Korpora

Real Academia Española - Corpus Diacrónico del Español (CORDE)

Consulta:	televisión				
Criterios de selección:					
Autor:			Obra:		
Cronológico:			Medio:	(Todos) Libros Periódicos Revistas Miscelánea Orales	Geográfico: (Todos) Argentina Bolivia Chile Colombia Costa Rica
Tema:	(Todos) 11.- Lírica 12.- Narrativa 121.- Breve 1211.- Relato breve tradicional 1212.- Relato breve culto				

Abbildung 19: CORDE – Suche nach *televisión* im gesamten Korpus.³⁴²

Mittels der genauen chronologischen, geographischen und genrespezifischen Einteilung der Texte ermöglicht das CORDE die Erstellung von Statistiken zu jedem Suchbegriff. Wenn der Nutzer beispielsweise nach dem Substantiv *televisión* im gesamten CORDE sucht, so werden 1047 Treffer in 96 Dokumenten gefunden. Nun kann er sich die Dokumente oder, nach weiterer Auswahl, die Konkordanzen anzeigen lassen, oder aber, er lässt eine Statistik erstellen:

Estadísticas								
Año	%	Casos	País	%	Casos	Tema	%	Casos
1965	51.11	526	ESPAÑA	91.18	952	15.- Prosa científica	57.21	599
1970	6.89	71	ARGENTINA	2.20	23	16.- Prosa de sociedad	14.80	155
1929	6.31	65	CUBA	1.53	16	12.- Prosa narrativa	11.46	120
1967	4.66	48	CHILE	1.34	14	19.- Prosa histórica	4.77	50
1971	4.56	47	ECUADOR	1.34	14	18.- Prosa periodística	4.68	49
1972	4.47	46	PERÚ	1.34	14	17.- Prosa religiosa	4.01	42
1974	3.59	37	MÉXICO	0.57	6	13.- Prosa dramática	1.33	14
1955	3.20	33	COLOMBIA	0.28	3	14.- Prosa didáctica	1.05	11
1947	2.52	26	EL SALVADOR	0.09	1	10.- Prosa jurídica	0.57	6
Otros	12.63	130	Otros	0.09	1	21.- Verso lírico	0.09	1

Abbildung 20: CORDE – Statistik der Ergebnisse des Substantivs *televisión*.³⁴³

³⁴² Real Academia Española (ed.): *CORDE*. Online Publikation: <http://corpus.rae.es/cordenet.html> (09.03.2017).

³⁴³ Real Academia Española (ed.): *CORDE. Estadística de 'televisión'*. Online Publikation: <http://corpus.rae.es/cgi-bin/crpsrvEx.dll> (09.03.2017).

Anhand dieser Statistik ist zu erkennen, dass das Substantiv *televisión* vor allem in Texten aus den 1970er und 1980er Jahren, insbesondere aus dem Jahr 1965, erscheint, auffallend viele Beispiele aus Spanien darunter sind (über 90% aller Belege), und diese vor allem in der wissenschaftlichen Prosa vorkommen.

2.2.1.2. CDH – *Corpus del Nuevo Diccionario Histórico del Español*

Ein weiteres historisches Korpus der *Real Academia Española* ist das *Corpus del Nuevo Diccionario Histórico del Español*. Das CDH umfasst 355 Mio. Lexeme in drei verschiedenen Subkorpora, die für Suchanfragen ausgewählt werden können: *Corpus nuclear del NDHE*, *S. XII-1975* und *1975-2000*.³⁴⁴

Das *Corpus nuclear del NDHE* beinhaltet mehr als 53 Mio. Eintragungen, von denen 32 Mio. aus spanischen Texten und mehr als 20 Mio. aus lateinamerikanischen Werken stammen. Die Texte, die das Korpus bilden, sind zum Großteil auch Bestandteil der Korpora CREA (siehe Kapitel 2.1.1.1.) und CORDE (siehe Kapitel 2.2.1.1.) und wurden semiautomatisch annotiert und lemmatisiert. Der Unterbereich *S. XII-1975* besteht aus Texten aus dem XII. Jahrhundert bis 1975, welche aus dem Korpus CORDE stammen und morphosyntaktisch annotiert sind. Es handelt sich hierbei um knapp 200 Mio. Formen. Das Subkorpus *1975-2000* umfasst mehr als 103 Mio. Lexeme, welche teilweise aus dem Korpus CREA stammen und ebenfalls linguistisch annotiert sind. Das CDH ist frei verfügbar und kostenlos im Internet zugänglich (<http://web.frl.es/CNDHE/view/inicioExterno.view>).³⁴⁵

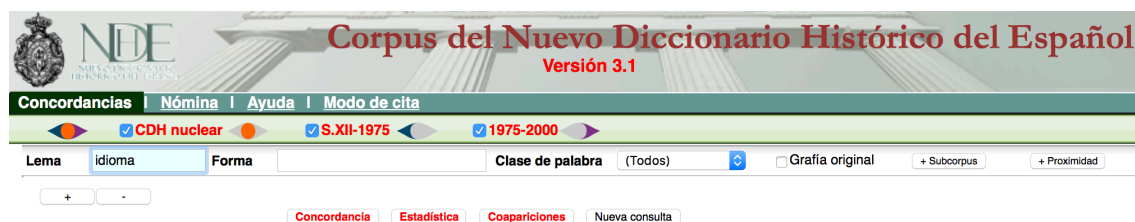


Abbildung 21: CDH – Suche nach dem Lemma *idioma*.³⁴⁶

Die Benutzeroberfläche des CDH ist ähnlich derjenigen des neuen, annotierten CREA und des CORPES aufgebaut. Es kann nach Lemmata und Formen unter Angabe der Wortart gesucht werden, wobei die Suche auf ein Subkorpus beschränkt und auch die

³⁴⁴ Vgl. Real Academia Española (ed.): *Banco de datos. CDH*. Online Publikation: <http://www.rae.es/recursos/banco-de-datos/cdh> (13.01.2017).

³⁴⁵ Vgl. Real Academia Española (ed.): *CDH*. Online Publikation: <http://web.frl.es/CNDHE/view/inicioExterno.view> (13.01.2017).

³⁴⁶ Real Academia Española (ed.): *CDH*. Op. cit. (17.04.2018).

2. Kategorisierung der Korpora

nähere Umgebung des Suchausdrucks über die Option *Proximidad* genauer bestimmt werden kann. Zusätzlich zur Anzeige von Konkordanzen und Statistiken gibt es beim CDH die Option *Coapariciones*, welche die Kollokationen der Suchbegriffe anzeigt (siehe Kapitel 1.4.8. und 4.1.1.).

	Clase	f	MI	LL SIMPLE	T-SCORE
guasteco	adjetivo	12	15.37	106.89	3.46
guasteco	sustantivo	12	15.37	106.89	3.46
vernáculo	sustantivo	15	13.37	109.17	3.87
neolatino	adjetivo	16	13.01	112.58	4.0
vernáculo	adjetivo	25	11.65	154.45	5.0
dialecto	sustantivo	73	11.47	442.98	8.54
otomí	adjetivo	18	11.43	108.69	4.24
otomí	sustantivo	18	11.28	107.02	4.24
lacio	sustantivo	16	11.2	94.37	4.0
indoeuropeo	adjetivo	13	11.14	76.19	3.6
vascuence	adjetivo	20	10.76	112.61	4.47
guaraní	adjetivo	26	10.67	144.85	5.09
vascuence	sustantivo	19	10.58	104.79	4.35
tagalo	adjetivo	15	10.43	81.43	3.87
tagalo	sustantivo	15	10.43	81.38	3.87
quechua	adjetivo	11	10.4	59.51	3.31
fonética	sustantivo	16	10.36	86.11	4.0
redactado	sustantivo	10	10.18	52.76	3.16
guaraní	sustantivo	23	10.12	120.42	4.79
fonético	adjetivo	22	10.09	114.77	4.69

1 - 20 of 1162

página: 1 2 3 4 5 6 7 8 9 ... 59

Abbildung 22: CDH – Ergebnis der Suche nach Kollokationen mit dem Substantiv *idioma*.³⁴⁷

Wenn man im gesamten Korpus nach dem Substantiv *idioma* sucht, erscheinen die obigen Kollokatoren, bei denen es sich meist um Adjektive und Substantive handelt. Zusätzlich zum jeweiligen Kollokator und dessen Wortart werden die absolute Häufigkeit (*f*) sowie die drei Werte der bereits in Kapitel 1.4.8. beschriebenen statistischen Verfahren zur Bestimmung von Kollokationen – *MI*, *LL-Simple* und *T-Score* – angegeben.

2.2.1.3. CdE:Orig (Gesamt) – *Corpus del Español (Género/Histórico)*

Wie bereits in Kapitel 2.1.1.3. erwähnt, besteht das CdE seit dem Sommer 2016 aus zwei verschiedenen Teilen.³⁴⁸ Das ursprüngliche CdE, welches nun CdE:Orig heißt, wurde im Jahr 2002 online veröffentlicht (www.corpusdelespanol.org) und umfasst als CdE:Orig (Gesamt) 100 Mio. Lexeme die aus über 20.000 Texten vom 13. bis 20. Jahrhundert stammen.³⁴⁹

³⁴⁷ Real Academia Española (ed.): *CDH. Entrada Compleja*. Online Publikation: <http://web.frl.es/CNDHE/org/publico/pages/consulta/entradaCompleja.view> (17.04.2018).

³⁴⁸ Vgl. Davies: *Corpus del Español. Corpora*. Op. cit. (21.11.2016).

³⁴⁹ Vgl. Davies: *Corpus del Español. Introducción*. Op. cit. (20.05.2011).

Während im CdE:Nuevo, welches im Bereich ‚Größe‘ in Kapitel 2.6.1. genauer beschrieben wird, vor allem das Internet und die verschiedenen Dialekte im Vordergrund stehen, können im CdE:Orig (Gesamt) verschiedene Genres und Jahrhunderte verglichen werden.³⁵⁰ Das Korpus ist lemmatisiert und nach Wortarten getaggt. Diese Tatsache erlaubt es beispielsweise, die verschiedenen Wortformen eines Lemmas zusammen zu gruppieren und die relative Häufigkeit desselben im Korpus anzugeben.³⁵¹

In Bezug auf die Aufteilung der Texte nach Epochen ist zu sagen, dass über 18 Mio. Lexeme aus dem Mittelalter stammen, mehr als 41 Mio. aus den *Siglos de Oro* und aus der zeitgenössischen Epoche ebenfalls über 41 Mio., wobei hier die Hälfte der Lexeme Teil des ausgeglichenen Subkorpus des aktuellen Spanisch ist, welches bereits als Referenzkorpus in Kapitel 2.1.1.3. vorgestellt wurde.³⁵²

Im Bereich der historischen Korpora werden das CdE:Orig (Gesamt) und das CORDE zwangsläufig miteinander verglichen. Das CORDE wurde von der *Real Academia Española* entwickelt, im Gegensatz dazu wurde das CdE:Orig (Gesamt) von einer einzigen Person, Mark Davies, erstellt. Genau hier sind aber auch einige Nachteile zu sehen, da das CdE:Orig (Gesamt) teilweise aus philologischer Sicht ungenau ist. Wegen der großen Datenmenge konnten die Texte nicht detailliert geprüft werden, weshalb sie immer wieder Fehler enthalten. Beim CORDE ist die Textgenauigkeit wesentlich höher, dafür ist die Bedienung nicht so komfortabel wie beim CdE:Orig (Gesamt), welches über einige nützliche Elemente verfügt, wie die grafische Darstellung bestimmter Entwicklungen oder die zielgenaue Suche durch die POS-getaggten Texte. Beide Korpora erlauben jedoch keinen Zugriff auf die kompletten Texte. Zudem bestehen beide aus zwei Komponenten, einer Konkordanzfunktion und einer Suchmaschine, mit der nach bestimmten Ausdrücken gesucht werden kann. Diese Suchmaschine verfügt bei beiden Korpora über eine Differenzierung nach Jahrhunderten, wobei das CORDE genauer nach Texttypen unterscheidet, das CdE:Orig (Gesamt) dafür aber die Suche nach grammatischen Kategorien zulässt. Das Suchergebnis umfasst die Anzeige der jeweiligen Form und die relative Häufigkeit derselben in den Texten. Durch die unterschiedliche

³⁵⁰ Vgl. Davies: *Corpus del Español – Género/Histórico*. Op. cit. (19.12.2016).

³⁵¹ Vgl. Davies/ Face (2006): op. cit., S. 134.

³⁵² Vgl. Davies, Mark: *Corpus del Español. Ayuda: Textos en el corpus*. Online Publikation: <http://www.corpusdelespanol.org/x.asp> (28.05.2011).

2. Kategorisierung der Korpora

Schwerpunktsetzung ergänzen sich beide Korpora und können so vorteilhaft genutzt werden.³⁵³

Ein Vorteil des CdE:Orig (Gesamt) ist demnach die grafische Darstellung der Nutzung von Lexemen in unterschiedlichen zeitlichen und genrespezifischen Bereichen:

Abbildung 23: CdE:Orig (Gesamt) – Benutzeroberfläche bei der Suche nach dem Substantiv *teléfono*.³⁵⁴

Wenn man beispielsweise im CdE:Orig (Gesamt) ohne Einschränkungen, also in allen zeitlichen und genrespezifischen Bereichen, nach einem Lexem wie *teléfono* sucht, so erhält man das folgende Ergebnis:

CONTEXT	ALL	1200s	1300s	1400s	1500s	1600s	1700s	1800s	1900s	ACAD	NEWS	FICT	ORAL
TELÉFONO	1363							30	1333	67	291	606	369

1.978 seconds

Abbildung 24: CdE:Orig (Gesamt) – Übersicht über das Erscheinen des Substantivs *teléfono*.³⁵⁵

Im CdE:Orig (Gesamt) erscheint *teléfono* also 30 Mal im 19. Jahrhundert und 1333 Mal im 20. Jahrhundert, was nicht weiter erstaunt, angesichts der Tatsache, dass das Telefon im Jahr 1861 erfunden wurde. Für das 20. Jahrhundert wird zudem die Aufteilung nach Genres angezeigt, wobei zu erkennen ist, dass das Substantiv *teléfono* vor allem im Bereich Fiktion, zusätzlich einige Male in mündlichen Texten und Nachrichten, aber nur selten in akademischen Texten erscheint.

2.2.1.4. Biblia Medieval

Das an der *Universitat de les Illes Balears* beheimatete Projekt der *Biblia Medieval* umfasst verschiedene mittelalterliche Übersetzungen der Bibel ins Kastilische. Es ist im Internet kostenlos und ohne Anmeldung zugänglich (<http://corpus.bibliamedieval.es>), um

³⁵³ Vgl. Kabatek/ Pusch (2011): op. cit., S. 216.

³⁵⁴ Davies: *Corpus del Español – Género/Histórico*. Op. cit. (09.03.2017).

³⁵⁵ Davies: *Corpus del Español – Género/Histórico*. Op. cit. (09.03.2017).

das Studium und die Verbreitung der mittelalterlichen Übersetzungen der Bibel ins Kastilische zu erleichtern. Das Korpus umfasst knapp fünf Mio. romanische und mehr als eineinhalb Mio. lateinische und hebräische Lexeme.³⁵⁶ Die Internetseite verfügt über einen Index derjenigen Manuskripte, die Eingang in das Korpus gefunden haben, mit Informationen zum Inhalt, wichtigen Daten zu den Texten, der ursprünglichen Sprache des Manuskripts und teilweise sogar Bildern der Schriften. Neben dem Zugang zum eigentlichen Korpus ist auch eine Seite mit Ressourcen verfügbar, die für das Studium der Bibelübersetzungen herangezogen werden können.³⁵⁷

In der Eingabemaske können die zu durchsuchenden Manuskripte ausgewählt und es kann speziell nach Kapiteln und Versen in jedem einzelnen Buch der Bibel gesucht werden. Die gefundenen Stellen werden daraufhin parallel dargestellt:³⁵⁸

Versículo	GE	E3	E4	ARRAGEL
GÉ 1:0	[fol. 1r-a] <i>Aqui se comiença la general & grand es toria que el muy noble Rey don alfonso fijo del noble Rey don fernando & dela Reyna donna beatriz mando fazer.</i>	[fol. 1r-a] ¶ <i>libro primero dela briuja que es llamado genesis & en ebrayco bere xid enel qual ay xij capitulos </i>	[fol. 1r-ab] <i>GENESIS EN EL NOMBRE DE DIOS Aqui comiença el primero libro dela bliuia, el qual es llamado en ebrayco beressit en latin genesi.</i>	[fol. 26r-a] <i>aqui comiença el libro prime ro dela ley en latin genesy es llamado & en ebrayco es llama do el libro de baresith: </i>
GÉ 1:1	[fol. 1r-b] <i> ... Delas obras que dios fizo en los prime ros .vj. dias. I Quando nuestro sennor dios crio enel comienço el çielo et la tierra ...</i>	<i>Capitulo primero de commo dios fizo el çielo & la tierra & de adan & de su liñaje fasta noe Enel co mien ço crio dios los çielos & la ti erra</i>	<i>Capitulo primero en que dizr commo dios crio el çielo la tierra dela obra que fizo enlos seys dias. & el dia septimo folgo. EN COMIENÇO Crio dios alos çielos & ala tierra</i>	<i>En el prinçipio crio el señor los çielos & la tierra </i>
GÉ 1:2		<i>& la ti erra era va na & vazia & la escuridat sobre la faz del abismo & espiritu de dios ventiscaua sobre fazes delas aguas</i>	<i>& la tierra era vana & vazia, & tiniebra sobre la faz del abismo. & el spiritu de dios auentaui sobre la faz delas aguas.</i>	<i>¶ & la tierra era vana & vazia & teneb ra sobre fazes del abismo E el spiritu de del señor era rretraydo sob re fazes delas aguas</i>

Abbildung 25: *Biblia Medieval* – Parallele Darstellung einiger Schriften zum ersten Kapitel des Buchs Genesis – Abschnitt I.³⁵⁹

³⁵⁶ Vgl. Enrique-Arias, Andrés/ Pueyo Mena, F. Javier (2008): *Corpus Biblia Medieval. Contenidos*. Online Publikation: <http://www.bibliamedieval.es/index.php/corpus-contenidos> (20.12.2016).

³⁵⁷ Vgl. Enrique-Arias, Andrés/ Pueyo Mena, F. Javier (2008): *Corpus Biblia Medieval. Página Principal*. Online Publikation: <http://www.bibliamedieval.es> (27.10.2016).

³⁵⁸ Anmerkung: Im Korpus werden alle Versionen nebeneinander dargestellt, doch aus Gründen der Lesbarkeit wurde die Anzeige hier in zwei Abschnitte aufgeteilt.

³⁵⁹ Enrique-Arias, Andrés/ Pueyo Mena, F. Javier (2008): *Corpus Biblia Medieval. Buscar primer capítulo de Génesis*. Online Publikation: <http://corpus.bibliamedieval.es> (09.03.2017).

2. Kategorisierung der Korpora

XV	LATIN	HEB	HEB-LEMA	HEB-TRANS
<p>Titulo de como Dios crio el cielo e la tierra e la mar e las arenas e las otras cosas que son vesibles e non vesibles. En el principio crio Dios los cielos e la tierra</p>	<p>in principio creavit Deus caelum et terram</p>	<p>בְּרֵאשִׁית בָּרָא אֱלֹהִים אֶת הַשָּׁמַיִם וְאֶת הָאָרֶץ:</p>	<p>בְּרֵאשִׁית בָּרָא אֱלֹהִים אֶת הַשָּׁמַיִם וְאֶת הָאָרֶץ</p>	<p>bere'sit bará' 'elohim 'et hašamayim ve'et ha'areš</p>
<p>vana e vazia e teniebra sobre fazes del abismo e el espiritu del Señor era retraido sobre fazes de las aguas.</p>	<p>terra autem erat inanis et vacua et tenebrae super faciem abyssi et spiritus Dei ferebatur super aquas</p>	<p>וְהָאָרֶץ הָיְתָה תֹהוּ וָבֹהוּ וַחֲשֹךְ עַל־פְּנֵי תְהוֹם וָרוּחַ אֱלֹהִים מְרַחֶפֶת עַל־פְּנֵי הַמַּיִם:</p>	<p>וְהָאָרֶץ הָיְתָה תֹהוּ וָבֹהוּ וַחֲשֹךְ עַל פְּנֵי תְהוֹם וָרוּחַ אֱלֹהִים רַח עַל פְּנֵי הַמַּיִם</p>	<p>veha'areš haytá tohu vabohu vehóšef 'al pené tehom verúah 'elohim meraḥéfet 'al pené hamáyim</p>

Abbildung 26: *Biblia Medieval* – Parallele Darstellung einiger Schriften zum ersten Kapitel des Buchs Genesis – Abschnitt II.³⁶⁰

Die mittelalterlichen kastilischen Übersetzungen der Bibel werden in diesem Korpus zusammen mit der hebräischen und lateinischen Originalversion gezeigt, wobei es teilweise auch möglich ist, Bilder des Originals zu sehen. Das Korpus ist damit als Parallelkorpus nutzbar, weshalb es besonders für Sprachhistoriker von Interesse sein dürfte.

2.2.2. Lernerkorpora

Neben den historischen Korpora präsentieren die Lernerkorpora eine andere Art der Funktionalität. Sie dienen der Sprachvermittlung und sollen Untersuchungen zu Problemen beim Spracherwerb dokumentieren. McEnery/Xiao unterscheiden auf dem Gebiet der unterrichtsorientierten Korpora das Lernerkorpus vom Entwicklungskorpus und dem LSP-Korpus (*Languages for Specific Purposes*). Während das sogenannte Lernerkorpus Sprachdaten von Lernern einer Zweit- oder Fremdsprache enthält, sind es beim Entwicklungskorpus Daten von Muttersprachlern und beim LSP-Korpus

³⁶⁰ Enrique-Arias/ Pueyo Mena (2008): *Corpus Biblia Medieval. Buscar primer capítulo de Génesis*. Op. cit. (09.03.2017).

bereichsspezifische Sprachdaten, wie beispielsweise akademisches Englisch. Es handelt sich allerdings bei allen dieser Korpora um Daten zum Spracherwerb.³⁶¹ In der Literatur wird zwischen Zweit- und Fremdspracherwerb unterschieden, bisweilen werden die Begriffe jedoch auch als austauschbar angesehen. Unter Zweitsprache versteht man die zweite Sprache, die sich eine Person aneignet und welche der Muttersprache in ihrer sozialen und kommunikativen Bedeutung nachgeordnet ist.³⁶² Generell wird davon ausgegangen, dass der Zweitspracherwerb in natürlicher Umgebung, gewöhnlich im zielsprachigen Ausland, stattfindet, wohingegen eine Fremdsprache normalerweise in einem Klassenzimmer gelernt wird, also fernab von extensiven natürlichen Sprachgelegenheiten.³⁶³

Lemnitzer/Zinsmeister geben an, dass in Lernerkorpora Schüleräußerungen gesammelt und typische Fehler annotiert werden.³⁶⁴ Aus diesem Grund werden sie bisweilen auch als Fehlerkorpora bezeichnet. Nachdem die Fehleranalyse (*Error Analysis*) zunächst aufgrund fehlender Möglichkeiten der qualitativen und quantitativen Auswertung verworfen worden war, ist sie seit der Entwicklung der großen, digitalen Korpora wieder aufgegriffen worden. Diese neue Variante der Fehleranalyse nennt sich *Computer-aided Error Analysis (CEA)*. Während zuvor kritisiert wurde, dass die Ergebnisse der Fehleranalyse nicht repräsentativ seien, beheben die neuen Möglichkeiten der modernen Technologie genau diese Nachteile. Aufgrund der strikten Vorgaben beim Aufbau des Korpus können die einzelnen Texte miteinander verglichen werden. Die Angaben über das Alter, die Herkunft und die sprachlichen Fähigkeiten der Sprecher sowie die Informationen über die Herkunft und das Genre der Texte ermöglichen eine genaue Einteilung der Beiträge und erlauben einen Vergleich zwischen Texten aus verschiedenen Korpora oder Subkorpora. Des Weiteren können nicht nur die Fehler der Lerner festgestellt werden, sondern auch die zu häufige oder zu seltene Verwendung von Lexemen oder Strukturen der Fremdsprache im Vergleich zu Muttersprachlern (*Over-*

³⁶¹ Vgl. McEnery, Tony / Xiao, Richard (2010): „What Corpora Can Offer in Language Teaching and Learning“. In: Hinkel, E. (Hrsg.) (2010): *Handbook of Research in Second Language Teaching and Learning: Volume 2*. London/ New York: Routledge, S. 371. Online Publikation: http://www.lancaster.ac.uk/fass/projects/corpus/ZJU/xpapers/McEnery_Xiao_teaching.PDF (12.06.2017).

³⁶² Vgl. Barkowski, Hans (2007): „Zweitsprachunterricht“. In: Bausch, K.-R./ Christ, H./ Krumm, H.-J. (Hrsg.) (2007): *Handbuch Fremdsprachenunterricht*. Tübingen/ Basel: A. Francke, S. 157.

³⁶³ Vgl. Nerbonne, John: *Natural Language Processing in Computer-Assisted Language Learning*. Online Publikation: <http://www.let.rug.nl/nerbonne/papers/nlp-hndbk-call.pdf> (29.10.2017), S. 5.

³⁶⁴ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 189.

2. Kategorisierung der Korpora

beziehungsweise *Underuse*). Auch die Kritik an der zu ungenauen oder fehlerhaften Klassifikation der von Schülern gemachten Fehler kann in Bezug auf die *CEA* nicht mehr geäußert werden. Durch strikte theoriegeleitete und psychologisch motivierte Taxonomien kann diese Schwäche der traditionellen Fehleranalyse behoben werden.³⁶⁵

Mithilfe der Lernerkorpora können also die spezifischen Fehler der Lerner ermittelt werden. Bisher wurden Schulbücher und Übungsmaterialien auf dem Wissen der Autoren bezüglich der allgemeinen Probleme der Schüler aufgebaut, anhand von Lernerkorpora könnten nun jedoch die für deutsche Schüler schwierigsten Probleme ermittelt und darauf abgestimmte Übungen entwickelt werden.

Guy Aston führt aus, dass Lernerkorpora in Bezug auf die Erstellung geeigneter Materialien für den Fremdsprachenunterricht sehr hilfreich sein können. Seiner Ansicht nach ist es fraglich, ob die Analyse des Verhaltens von Muttersprachlern eine adäquate Grundlage für die Erstellung des Lehrplans ist, da dadurch keinerlei Auskunft über die relative Schwierigkeit und Lernbarkeit spezieller Eigenheiten der Sprache, die unterrichtet werden sollen, gegeben werden kann. Er vertritt die Meinung, dass relevante Informationen in Bezug auf zwischensprachliche Entwicklungen durch Lernerkorpora bereitgestellt werden können. Diese können verwendet werden, um wiederkehrende Fehler zusammen mit *Over-* und *Underuses* zu kodifizieren und zu klassifizieren, um diejenigen sprachlichen Eigenheiten zu identifizieren, die im Unterricht betont werden sollten, und um deren Schwierigkeitsgrad zu ermitteln. Außerdem berücksichtigen Materialien, die auf Daten von Muttersprachlern beruhen, nicht die Produktivität sprachlicher Charakteristika aus der Perspektive von Nicht-Muttersprachlern.³⁶⁶

Im Hinblick auf die spanische Sprache gibt es einige Lernerkorpora, welche im Folgenden vorgestellt werden sollen. Dabei handelt es sich um drei rein mündliche Korpora, zwei Korpora der geschriebenen Sprache sowie ein Korpus mit gesprochenen und geschriebenen Texten. Zur ersten Kategorie zählen das *Corpus Oral de Español como Lengua Extranjera (ELE)* (kurz: CORELE), die beiden Subkorpora des SPLLOC (*Spanish Learner Language Oral Corpora*) sowie das *Fono.ele*. Das *Corpus de*

³⁶⁵ Vgl. Leńko-Szymańska, Agnieszka (2003): „Lexical problem areas in the advanced learner corpus of written data“. In: Lewandowska-Tomaszczyk, B./ Melia, P.J. (Hrsg.) (2003): *PALC 2001: Practical Applications in Language Corpora*. Frankfurt am Main: Peter Lang, S. 507-508.

³⁶⁶ Vgl. Aston, Guy (2000): „Corpora and language teaching“. In Burnard, L./ McEnery, T. (Hrsg.) (2000): *Rethinking Language Pedagogy from a Corpus Perspective*. Frankfurt am Main: Peter Lang, S. 11.

Aprendices de Español como Lengua Extranjera (CAES) und das *Corpus Escrito del Español L2* (CEDEL2) enthalten im Gegensatz dazu ausschließlich geschriebene Daten und das LANGSNAP (*Languages and Social Networks Abroad*) ist ein Korpus der gesprochenen und geschriebenen Sprache.

2.2.2.1. CORELE – *Corpus Oral de Español como Lengua Extranjera* (ELE)

Das *Corpus Oral de Español como Lengua Extranjera* (ELE) wurde an der *Universidad Autónoma de Madrid* entwickelt und enthält 40 Interviews von Spanischlernern mit über 9 verschiedenen Muttersprachen. Die Mehrheit der Spanischler war zum Zeitpunkt der Interviews zwischen 18 und 26 Jahren alt und belegte Kurse der Niveaus A2 und B1 des *Common European Framework of Reference* (*Gemeinsamer Europäischer Referenzrahmen für Sprachen* (GER) – siehe auch Kapitel 3.1.2.). Die gesamten Aufnahmen der Lerner umfassen 55.567 Lexeme (13 Stunden und 36 Minuten Tonmaterial). Zudem wurde eine Kontrollgruppe von vier Muttersprachlern interviewt: Diese Interviews umfassen insgesamt 9.389 Lexeme (1 Stunde und 22 Minuten Tonmaterial). Die Dateien enthalten eine orthographische Transkription, Metadaten mit soziolinguistischer Information (z. B. Herkunft und Bildungsniveau) sowie Angaben zu den Spanischkenntnissen der Interviewten (z. B. Leistungsstand). Die Texte wurden zudem mit Fehlerannotation und morphosyntaktischer Annotation versehen.³⁶⁷ Das Korpus ist frei im Internet zugänglich (<http://cartago.llf.uam.es/corele/index.html>).³⁶⁸

Der Nutzer kann sich die Interviews entweder komplett anhören oder speziell nach Informationen suchen. Zunächst kann er auswählen, ob er nach Lemmata, Formen oder Wortarten suchen will (wobei zwei Optionen angegeben werden können) sowie, ob er die Dateien nach der Muttersprache des Interviewten oder deren Sprachniveau sortieren will. Daraufhin werden die Treffer untereinander aufgereiht, und der Nutzer kann sich die betreffenden Ausschnitte, die den Suchbegriff enthalten, anhören sowie Informationen über den Interviewten und die Häufigkeit des Begriffs anzeigen lassen. Zudem sind Listen der häufigsten Lexeme verfügbar und es werden die auffälligsten Aussprachephänomene für jede Muttersprache zusammen mit Textbeispielen (Transkription mit Hörbeispielen)

³⁶⁷ Vgl. Laboratorio de Lingüística Informática de la Universidad Autónoma de Madrid (ed.): *Spanish Learner Oral Corpus*. Online Publikation: <http://www.llf.uam.es/ING/SLOC.html> (11.01.2017).

³⁶⁸ Vgl. Universidad Autónoma de Madrid (ed.): *Corpus de Español como Lengua Extranjera* (ELE). Online Publikation: <http://cartago.llf.uam.es/corele/index.html> (12.01.2017).

2. Kategorisierung der Korpora

aufgelistet. Wie bei allen Sprachen in diesem Korpus handelt es sich in Bezug auf Lerner mit deutscher Muttersprache allerdings nur um vier Interviews.

Die Fehler der Interviewten kann sich der Nutzer entweder ebenfalls nach Muttersprachen sortiert anzeigen lassen oder nach Sprachbereich (Aussprache, Grammatik, Lexik oder Pragmatik), Wortart, Fehlerart oder Sprachniveau. Es können auch mehrere Optionen ausgewählt werden, sodass beispielsweise gezielt nach Fehlern deutscher Muttersprachler im lexikalischen Bereich gesucht werden kann. Jeder einzelne Fehler wird bei einem Klick darauf genau kategorisiert und die Verbesserung angezeigt. Im CORELE kann somit nach Fehlern deutscher Muttersprachler im Hinblick auf spanische Verben gesucht werden. In diesem Bereich werden folgende Fehler angezeigt:

Resultados de errores de verbo en estudiantes cuya lengua materna es el alemán:

THO: yo {%pho: ['jo]} creo que → &ah español es una importante {%pho: [ɪmpo:'tante]} idioma para /

ENT: hhh {%act: assent} ///

THO: → &ah todo {%pho: ['toɾo]} el mundo porque &ah mucha gente hablan en Suramérica español también {%pho: [dam'bjen]} ///

-Error: **hablan**

THO: conozco {%pho: [ko'nosko]} &ah alguna gente {%pho: ['hente]} de Suramérica y vive con → tres de ellos {%pho: ['ejos]} pero /

-Error: **vive**

ENT: hhh {%act: assent} / <pero no &ha [/] conoces el país> ///

THO: → [<] <nunca he ido / ahí {%pho: [a'e]} / no / sí> ///

Abbildung 27: CORELE – Analyse von Fehlern deutscher Muttersprachler im Hinblick auf Verbformen.³⁶⁹

Durch einen Klick auf den Pfeil am linken Rand kann der Nutzer sich die Audiodatei zum jeweiligen Sprachbeispiel anhören, das Symbol darunter führt zu genauen Informationen zum jeweiligen Sprecher und die Lupe neben der Fehleranzeige öffnet eine präzise Fehleranalyse mit Verbesserung.

³⁶⁹ Universidad Autónoma de Madrid (ed.): *Corpus de Español como Lengua Extranjera (ELE)*. *Resultados de errores de verbo en estudiantes cuya lengua materna es el alemán*. Online Publikation: http://cartago.llf.uam.es/exist/rest/db/corpus/L1_y_cat.xql?category=V&L1=German (10.03.2017).

2.2.2.2. SPLLOC – *Spanish Learner Language Oral Corpora*

Bei SPLLOC handelt es sich um zwei spanische Lernerkorpora, die an den Universitäten Southampton, Newcastle und York zusammengestellt wurden und im Internet kostenlos zur Verfügung stehen (<http://www.splloc.soton.ac.uk>).³⁷⁰

Das Ziel dieses Projekts ist es, die Aneignung des Spanischen als Fremdsprache im Unterricht zu untersuchen, um damit sowohl Rückschlüsse auf die Erstellung von Lehrplänen ziehen als auch fundamentale theoretische Fragen hinsichtlich des Zweit- und Fremdsprachenerwerbs klären zu können. Das Programm besteht aus zwei Forschungsprojekten SPLLOC1 (von 2006 bis 2008) und SPLLOC2 (von 2008 bis 2010). In beiden Projekten wurden Daten aus dem Spanischunterricht englischer Muttersprachler von Anfängern bis Fortgeschrittenen mittels speziell designter Aufgaben gesammelt. Zu Vergleichszwecken wurden auch Muttersprachler bei der Beantwortung der Aufgaben aufgezeichnet. Die in SPLLOC enthaltenen Texte sind unter anderem Erzählungen, Interviews, Bildbeschreibungen und Paardiskussionen.³⁷¹ Auf der Homepage sind über 500 Audiodateien sowie deren Transkripte verfügbar.³⁷²

Die Korpora erlauben Sucheinstellungen bezüglich der Aufgabenart, dem Sprachniveau des Lernalters, dessen Geschlecht, der Anzahl der Lernjahre sowie anderer gelernter Sprachen, sie bieten allerdings keine Möglichkeit der Suche nach Lexemen, Lemmata oder Wortarten.³⁷³

³⁷⁰ Vgl. Domínguez, Laura/ Mitchell, Rosamond/ Myles, Florence/ Marsden, Emma: *Spanish Learner Language Oral Corpora. Linguistic development in L2 Spanish. Home Page*. Online Publikation: <http://www.splloc.soton.ac.uk/index.html> (09.10.2017).

³⁷¹ Vgl. Domínguez, Laura/ Mitchell, Rosamond/ Myles, Florence/ Marsden, Emma: *Spanish Learner Language Oral Corpora. Linguistic development in L2 Spanish. Home Page*. Op. cit. (09.10.2017).

³⁷² Vgl. Domínguez, Laura/ Mitchell, Rosamond/ Myles, Florence/ Marsden, Emma: *Spanish Learner Language Oral Corpora. Linguistic development in L2 Spanish. Search Results*. Online Publikation: <http://www.splloc.soton.ac.uk/dosearch.php> (13.10.2017).

³⁷³ Vgl. Domínguez, Laura/ Mitchell, Rosamond/ Myles, Florence/ Marsden, Emma: *Spanish Learner Language Oral Corpora. Linguistic development in L2 Spanish. Search the SPLLOC Corpus*. Online Publikation: <http://www.splloc.soton.ac.uk/search.php> (12.10.2017).

2. Kategorisierung der Korpora

```
@UTF8
@Begin
@Languages:      es
@Participants:   H57 Participant, SMP Investigator
@ID:             es|splloc2|H57|male|Yaer13|Participant||
@ID:             es|splloc2|SMP|female||Investigator||
@Date:           30-JAN-2009
@Location:       PS
@Situation:      Las Hermanas Story
@Transcriber:    TIG
@Time Duration:  00:04:39
*SMP:            vale cuÃ©ntame la historia de las hermanas .
*H57:            es de dos hermanas que fueron de vacaciones en EspaÃ±a .
*H57:            en principio fueron a Madrid para visitar la ciudad y .
*H57:            fueron a [/] a [/] al estadio de fÃ©tbol y unos edificios .
*H57:            tambiÃ©n comÃ©an tapas y +/.
*H57:            pues no sÃ© los dos pero .
*H57:            +, estoy seguro que Sarah comÃ©a tapas y Gwen bebÃ©a vino y .
*H57:            despuÃ©s decidieron ir a Barcelona .
*H57:            por eso cogÃ©an el tren y (.) .
*H57:            entraban en el tren y .
*H57:            durante la [/] el viaje hablaban de [/] sobre su niÃ±ez y .
*H57:            su niÃ±ez um cuÃ©nto de pequeÃ±as eran muy diferentes en mil
*H57:            novecientos noventa y (.) seis .
*H57:            Gwen de niÃ±a cada semana fue muy +/.
*H57:            cÃ³mo se dice ?
```

Abbildung 28: SPLLOC – Transkript einer Geschichte zur Aufgabe *Dos Hermanas*.³⁷⁴

Die grundlegenden Prinzipien der Korpora sind Ausgeglichenheit hinsichtlich spontaner und gelenkter Sprache, hinsichtlich der Genres (Interview, Erzählung und Diskussion) und der Teilnehmer (20 Sprecher aus jedem der drei Levels, zusätzlich Muttersprachler), Flexibilität der computerbasierten Analyse sowie freier Zugriff für alle Interessierten auf alle Materialien.³⁷⁵

2.2.2.3. Fono.ele

Das Projekt *Adquisición y aprendizaje del componente fónico del español como lengua extranjera/segunda lengua* der *Universidad de Alcalá* erstellte das Korpus *Fono.ele* für Untersuchungen zur Phonetik des Spanischen als Fremd- oder Zweitsprache. Es handelt sich bei *Fono.ele* um ein mündliches Lernerkorpus. Die Homepage soll dem Austausch über Forschungsergebnisse dienen und nicht nur Wissenschaftlern, sondern auch Lehrern von Nutzen sein. Zudem soll dadurch die Vermittlung der phonischen

³⁷⁴ Domínguez, Laura/ Mitchell, Rosamond/ Myles, Florence/ Marsden, Emma: *Spanish Learner Language Oral Corpora. Linguistic development in L2 Spanish. Transcript Hermanas*. Online Publikation: <http://www.splloc.soton.ac.uk/data/splloc2/transcripts/Hermanas/H57SMP13.cha> (12.10.2017).

³⁷⁵ Vgl. Domínguez, Laura/ Mitchell, Rosamond/ Myles, Florence/ Marsden, Emma: *Spanish Learner Language Oral Corpora. Linguistic development in L2 Spanish. Rationale*. Online Publikation: <http://www.splloc.soton.ac.uk/rationale.html> (12.10.2017).

Komponente des Spanischen und damit die Lehre der spanischen Sprache im Allgemeinen vorangebracht werden.³⁷⁶

Derzeit verfügt das Korpus *Fono.ele* über Daten von Studenten sechs verschiedener Nationalitäten: Deutsch, Griechisch, Taiwanisch, Polnisch, Portugiesisch und Ägyptisch. Ziel des Korpus ist es, Forschungsvorhaben sowohl theoretischer als auch praktischer Art auf dem Gebiet der Aneignung der phonischen Komponente des Spanischen als Fremdsprache zu erleichtern. Im Hinblick darauf bietet es verschiedenste Sprachdaten, sowohl in soziokultureller als auch in linguistischer Hinsicht. Es soll außerdem Materialien für Lehrer des Spanischen als Fremdsprache sowie für fortgeschrittene Lerner zur Verfügung stellen, um deren Wissen zu erweitern.³⁷⁷

Buscar Audios

Tipo de texto:

Identificación del texto:

Contenido del texto:

Nº Informante:

Nacionalidad:

Sexo:

Grupo de edad:

Contacto con el Español:

Nivel de la lengua:

Experiencias de aprendizaje:

Mostrando: 1-6 de 6 audios

Tipo	Id. Texto	Nº Inf.	Nac.	Sexo	Edad	Cont. Esp.	Nivel	Exp. Apren.	mp3
Conversación	conversacion_nivel_A	L01	Alemania	Mujer	Grupo 1	bajo	A2	si ha trabajado la fonética	▶
Conversación	conversacion_nivel_A	L02	Alemania	Mujer	Grupo 1	bajo	A2	algunas veces	▶
Conversación	conversacion_nivel_B	L07	Alemania	Mujer	Grupo 1	bajo	B1	algunas veces	▶
Conversación	conversacion_nivel_B	L08	Alemania	Hombre	Grupo 1	muy bajo	B1	si ha trabajado la fonética	▶
Conversación	conversacion_nivel_C	L15	Alemania	Hombre	Grupo 1	medio	C1	algunas veces	▶
Conversación	conversacion_nivel_C	L16	Alemania	Mujer	Grupo 1	medio	C1	algunas veces	▶

Abbildung 29: *Fono.ele* – Konversationen von deutschen Spanischlernern.³⁷⁸

Die Benutzeroberfläche erlaubt Einstellungen zur Art des Texts (Konversation, Text, Satz oder Lexem), zur Nationalität, zum Geschlecht, zur Altersgruppe (*grupo 1*: 18-25 Jahre; *grupo 2*: 26-35 Jahre), zum Umfang des Kontakts mit der spanischen Sprache, zum Sprachniveau (von A2 bis C1) sowie zu den Lernerfahrungen im Bereich der Phonetik.³⁷⁹ Das Korpus *Fono.ele* ist im Internet kostenlos zugänglich

³⁷⁶ Vgl. Blanco Canales, Ana/ Álvarez Martínez, María Ángeles: *Corpus Fono.ele. Presentación*. Online Publikation: <http://www3.uah.es/fonoele/index.php> (13.10.2017).

³⁷⁷ Vgl. Blanco Canales, Ana/ Álvarez Martínez, María Ángeles: *Corpus Fono.ele. Descripción, objetivos, acceso*. Online Publikation: <http://www3.uah.es/fonoele/corpus.php> (13.10.2017).

³⁷⁸ Blanco Canales, Ana/ Álvarez Martínez, María Ángeles: *Corpus Fono.ele. Acceso al corpus*. Online Publikation: <http://www3.uah.es/fonoele/> (13.10.2017).

³⁷⁹ Vgl. Blanco Canales, Ana/ Álvarez Martínez, María Ángeles: *Corpus Fono.ele. Acceso al corpus*. op cit. (13.10.2017).

(<http://www3.uah.es/fonoele/corpus-muestra.php>) und stellt 630 Audiodateien zur Verfügung, allerdings keinerlei Transkripte.

2.2.2.4. LANGSNAP – *Languages and Social Networks Abroad*

LANGSNAP ist ein Lernerkorpus mit Texten von englischen Muttersprachlern, das an der Universität Southampton erstellt wurde. Es enthält mündliche und schriftliche Texte von Spanisch- und Französischlernern, die das dritte Jahr ihres vierjährigen Studiums als Studenten, *teaching assistants* oder Praktikanten in Frankreich, Spanien oder Mexiko verbrachten. Die Lernerkorpora wurden in den Jahren 2011 bis 2013 während des Projekts *Social networks, target language interaction, and second language acquisition during the year abroad: A longitudinal study* gesammelt. Von jedem Teilnehmer wurden insgesamt sechs Mal Daten erfasst – vor, während und nach dem Auslandsaufenthalt. Bei den erstellten Texten handelt es sich um mündliche Interviews, die Wiedergabe einer Bildergeschichte und das Verfassen eines argumentativen Aufsatzes.³⁸⁰

Das Ziel des LANGSNAP Projekts war es, die Entwicklung des Wissens über und des Gebrauchs der Fremdsprachen Spanisch und Französisch von fortgeschrittenen Studenten während eines Zeitraums von 21 Monaten inklusive eines neunmonatigen Auslandsaufenthalts zu dokumentieren und der Forschungsgemeinschaft zugänglich zu machen. Von besonderem Interesse waren die Aneignung der jeweiligen Fremdsprache während des Auslandsaufenthalts, die Entwicklung sozialer Netzwerke während dieser Zeit, die Art und das Ausmaß der Einflussfaktoren auf den Sprachgebrauch im Ausland, die Arten der Lerngelegenheiten durch die zielsprachliche Interaktion während des Auslandsaufenthalts sowie die Beziehungen zwischen dem Aufbau sozialer Netzwerke, der sozialen Interaktion und dem Spracherwerb. Es wurden Daten von 56 englischsprachigen Universitätsstudenten sowie zu Vergleichszwecken von zehn französischen und spanischen Muttersprachlern erhoben. Die Daten können nach Teilnehmer, Aktivität oder Erhebungsrunde (vor, während oder nach dem Auslandsaufenthalt) sortiert werden.³⁸¹

³⁸⁰ Vgl. Mitchell, Rosamond/ Richard, Laurence/ Romero de Mills, Patricia: *Welcome to LANGSNAP*. Online Publikation: <http://langsnap.soton.ac.uk> (12.10.2017).

³⁸¹ Vgl. Mitchell, Rosamond/ Richard, Laurence/ Romero de Mills, Patricia: *The LANGSNAP Project: Languages and Social Networks Abroad*. Online Publikation: <http://langsnap.soton.ac.uk/project.html> (12.10.2017).

*150: le gente sigue fumando y bebiendo al exceso sin pensar en los
costes y otros efectos negativos .
%mor: pro:per|le=him n|gente&fem=folk v|seguir-3S&PRES=follow
ger|fuma-PROG=smoke coord|y=and ger|bebe-PROG=drink
aprep|a-det|el&MASC n|exceso&masc=excess prep|sin=without
inf|piensa-INF=think prep|en=in det:art|el&MASC-PL=the
v|costa-2S&SUB&PRES=cost coord|y=and det:indef|otro-MASC-PL=other
n|efecto&masc-PL=effect adj|negativo-MASC-PL=negative .

Abbildung 30: LANGSNAP – Transkript eines argumentativen Aufsatzes zum Thema *Drugs* inklusive Grammatikanalyse.³⁸²

Auf der Homepage sind neben den Audiodateien und bloßen Transkripten auch Transkripte mit Grammatikanalysen verfügbar, wie in der obigen Abbildung zu sehen ist. Die Dateien sind im Internet kostenlos und frei zugänglich (<http://langsnap.soton.ac.uk>).³⁸³

2.2.2.5. CAES – *Corpus de aprendices de español como lengua extranjera*

Das *Corpus de Aprendices de Español como Lengua Extranjera* (kurz: CAES) wurde vom *Instituto Cervantes* sowohl gefördert als auch finanziert, und von einem Forschungsteam der *Universidad de Santiago de Compostela* erstellt. Es handelt sich dabei um eine Sammlung von Texten, welche von Spanischstudenten mit unterschiedlichen Sprachniveaus des GER von A1 bis C1 und sechs verschiedenen Muttersprachen (Arabisch, Chinesisch, Französisch, Englisch, Portugiesisch und Russisch) in den Jahren 2011 bis 2013 verfasst wurden. Das Korpus CAES soll unter anderem für Lehrer, Forscher, Sprachtester und Verleger von Unterrichtsmaterialien eine solide und objektive Quelle für Daten hinsichtlich Schwierigkeiten beim Fremdsprachenerwerb, gewöhnlicher Fehlerquellen oder genutztem Vokabular darstellen und ist kostenlos im Internet zugänglich (<http://galvan.usc.es/caes/search>). Derzeit umfasst es 575.000 Lexeme aus Texten von 1.423 Studenten, welche dafür jeweils zwei bis drei Texte verfassten. Die sich daraus ergebenden 3.878 Texte wurden automatisch morphosyntaktisch annotiert und später manuell disambiguiert.³⁸⁴

³⁸² Mitchell, Rosamond/ Richard, Laurence/ Romero de Mills, Patricia: *The LANGSNAP Project: Writing (Drugs) – Participant 150. Text (Transcript with Grammar Analysis)*. Online Publikation: <http://langsnap.soton.ac.uk/850/2/F150c.mor.pst.cex> (12.10.2017).

³⁸³ Vgl. Mitchell, Rosamond/ Richard, Laurence/ Romero de Mills, Patricia: *The LANGSNAP Project: Writing (Drugs) – Participant 150. Full Record*. Online Publikation: <http://langsnap.soton.ac.uk/849/> (12.10.2017).

³⁸⁴ Vgl. Instituto Cervantes (ed.): *Corpus de aprendices de español (CAES)*. Online Publikation: http://www.cervantes.es/lengua_y_ensenanza/tecnologia_espanol/caes.htm (15.10.2017).

2. Kategorisierung der Korpora

Im CAES können Sucheinstellungen zu Sprachniveau, Muttersprache, Herkunft, Alter und Geschlecht des Lernalters vorgenommen werden. Zudem kann nach Lemmata und Wortarten gesucht sowie zwischen Schreibungen mit und ohne Akzent sowie Groß- und Kleinschreibung unterschieden werden.³⁸⁵

personas de todas las edades , los jóvenes haciendo	ejercicio , los niños jugando , todos pasando una buena tarde .
hago ballet desde niña y ningun otro tipo de	ejercicio , me disgusta mucho el gimnasio , a mí me parece muy monótono
en la playa con mi perro por hacer de	ejercicio .
Es muy importante que tenga una gimnasio para hacer	ejercicio .
esposa le gusta nadar y yo me gusta hacer	ejercicio .
Ella es muy saludable porque hacer mucho	ejercicio .
Paige es muy divertida y le gusta hacer	ejercicio en el gimnasio , va a deportes con tus amigos y compra en
Es un bueno	ejercicio en la evolución de los personajes en un mundo típico y remoto .
, también porque era un hombre que hacia mucho	ejercicio físico .
Acaso puede ser que algunos colegas podrían compartir este	ejercicio pero con discreción porque la competición está muy fuerte en este tema .
Pues , ella esta en buen forma y hace	ejercicio tres veces por semana .
Podríamos hacer	ejercicio y adoptar formas de reducir el estres o buscar tratamiento mas natural .
, hacer camping , ir de compras , hacer	ejercicio y escuchar a música .
, tengo 2 hermanos , gosto mucho de hacer	ejercicio y hablo en inglés .
el lugar mas cerca de mi casa para hacer	ejercicio .

Abbildung 31: CAES – Konkordanzen des Substantivs *ejercicio*.³⁸⁶

Die Ergebnisse können als Konkordanzliste oder als Statistik angezeigt werden. In Letzterer wird ersichtlich, bei welchen Sprechern – also auf welchem sprachlichen Niveau, mit welcher Muttersprache und mit welchem Geschlecht – der Begriff verwendet wird.³⁸⁷

En Paris	tengo	una	tía	que	trabaja	cerca de la	Catedral_de_Notre_Dame_de_Paris ,
X	NPG	VMIP1S	DA-FS	NCFS	PL+CN	VMUP9S	WP+ X DA+FS NPG P
en París	tener	un	tío	que	trabajar	cerca de el	Catedral_de_Notre_Dame_de_Paris ,

Abbildung 32: CAES – Substantiv *ejercicio* innerhalb des Texts einer Portugiesin mit dem Sprachniveau A2.³⁸⁸

³⁸⁵ Vgl. Instituto Cervantes: *Corpus de aprendices de español (CAES)*. *Buscar*. Online Publikation: <http://galvan.usc.es/caes/search> (15.10.2017).

³⁸⁶ Instituto Cervantes: *Corpus de aprendices de español (CAES)*. *Buscar ejercicio. Ejemplos*. Online Publikation: http://galvan.usc.es/caes/search?utf8=✓&level=Cualquiera&mother_tongue=Cualquiera&country=Cualquiera&age_from=&age_to=&sex=Cualquiera&distance=&token1=ejercicio&tag1=&lemma1=&token2=&tag2=&lemma2=&token3=&tag3=&lemma3=&token4=&tag4=&lemma4=&result_type=Ejemplos&ordering=Elemento&page_size=50&commit=Buscar (15.10.2017).

³⁸⁷ Vgl. Instituto Cervantes: *Corpus de aprendices de español (CAES)*. *Buscar ejercicio. Estadística*. Online Publikation: http://galvan.usc.es/caes/search?utf8=✓&level=Cualquiera&mother_tongue=Cualquiera&country=Cualquiera&age_from=&age_to=&sex=Cualquiera&distance=&token1=ejercicio&tag1=&lemma1=&token2=&tag2=&lemma2=&token3=&tag3=&lemma3=&token4=&tag4=&lemma4=&result_type=Estad%C3%ADstica+completa&commit=Buscar (15.10.2017).

³⁸⁸ Instituto Cervantes: *Corpus de aprendices de español (CAES)*. *Buscar ejercicio. Contexto*. Online Publikation: <http://galvan.usc.es/caes/tokens/m3by4yaus279ppo3?lemma1=&lemma2=&lemma3=&lemma4=&tag1=&tag2=&tag3=&tag4=&token1=ejercicio&token2=&token3=&token4=> (15.10.2017).

Des Weiteren kann sich der Nutzer jedes Beispiel im Kontext anzeigen lassen, wie in obigem Beispiel zu sehen ist. Dabei werden die morphosyntaktische Annotation sowie das jeweilige Lemma angezeigt.

2.2.2.6. CEDEL2 – *Corpus Escrito del Español L2*

Das *Corpus Escrito del Español L2* (kurz: CEDEL2) ist ein spanisches Lernerkorpus aus Texten von Spanischlernern mit englischer Muttersprache, wobei die Mehrheit der Lerner aus den USA stammt. Neben Texten englischer Spanischlerner verfügt das Korpus zu Vergleichszwecken auch über eine große Anzahl von Aufsätzen spanischer Muttersprachler. Es enthält derzeit mehr als 800.000 Lexeme von über 2.500 Teilnehmern, darunter über 1.600 englischsprachige Spanischlerner und fast 800 spanische Muttersprachler.³⁸⁹

CEDEL2 wurde von Cristóbal Lozano erstellt und ist Teil des WOSLAC Projekts (*Word Order in Second Language Acquisition Corpora*) unter der Leitung von Amaya Mendikoetxea von der *Universidad Autónoma de Madrid*. Anhand des Korpus soll untersucht werden, wie englischsprachige Spanischlerner die spanische Grammatik erlernen. Das Korpus befindet sich noch im Aufbau und soll in seiner endgültigen Fassung mit einer Mio. Lexemen Forschern, Lehrern und Spanischlernern dienlich sein. Um dieses Ziel zu erreichen werden fortlaufend Spanischlerner und Muttersprachler zur Teilnahme an diesem Projekt geworben und ihre Aufsätze gesammelt. Das Korpus soll nach der Fertigstellung allen Interessierten zugänglich gemacht werden, doch ein Teil desselben ist bereits im Internet verfügbar (<http://cedel2.learnercorpora.com>). Zu Vergleichszwecken wird außerdem ein Korpus namens WriCLE (*Written Corpus of Learner English*) mit Daten von spanischen Muttersprachlern, die Englisch lernen, erstellt.³⁹⁰

In CEDEL2 sind derzeit drei Subkorpora verfügbar: *Spanish natives*, *L1 English – L2 Spanish* und *L1 Greek – L2 Spanish*. In den Lernerkorpora können die Texte nach dem Alter und Sprachniveau der Lerner, der Dauer des bisherigen Spanischerwerbs, eventuellen Auslandsaufenthalten, weiteren Fremdsprachen sowie dem Aufsatzthema ausgewählt werden. Die Texte der Muttersprachler werden unter Angabe des

³⁸⁹ Vgl. Lozano, Cristóbal: *CEDEL2 info*. Online Publikation: <http://www.uam.es/proyectosinv/woslac/collaborating.htm> (11.07.2019).

³⁹⁰ Vgl. Lozano: *CEDEL2 info*. Op. cit. (16.04.2018).

2. Kategorisierung der Korpora

Geschlechts, des Alters, des Geburtsorts und der gelernten Fremdsprachen des Sprechers sowie nach dem Thema des Aufsatzes kategorisiert. Es stehen unter anderem Texte zu Themen wie *¿Cómo es la región donde vives?* / *Habla de una persona famosa.* / *Resume una película que has visto recientemente.* oder *¿Qué hiciste el año pasado durante las vacaciones?* zur Auswahl. Neben diesen alltäglichen Themen sind allerdings auch Aufsätze zu drängenden Problemen wie Terrorismus und Einwanderung verfasst worden.

Anhand der Benutzeroberfläche kann nach Dateinamen oder einzelnen Lexemen gesucht werden, wie beispielsweise dem Substantiv *paisaje* im folgenden Aufsatz einer 61-jährigen Spanierin zu ihrer Heimat. Zusätzlich zum Text und dem jeweils farblich markierten Suchbegriff werden alle Informationen zur jeweiligen Datei angegeben.

01. ¿Cómo es la región donde vives?

La región dónde vivo es Catalunya. Es un país pequeño pero precioso que goza de una gente, **paisaje**, clima y cultura muy diversa y muy interesante. Está situado al noroeste de España y hace frontera con Francia por lo que su conexión con Europa es fácil. Lo que más me gusta de su geografía es que tiene altas montañas: los Pirineos; un gran río: el Ebro y también una gran costa mediterránea, además de bosques y diferentes tierras de cultivo, lo que convierte a Catalunya en una región rica y diversa. Posee idioma propio, el catalán, que está muy extendido y se habla junto con el español. Su cultura es muy rica y los catalanes se sienten orgullosos de ella y procuran mantenerla utilizando todos los recursos que poseen.

Participant's data

ID	es_wr_61_1_MDA
Participant type	Native
Native Language	Spanish
Country	España
Sex	Female
Age	61
Foreign languages	
Composition theme	01. ¿Cómo es la región donde vives?
Word count	146

Abbildung 33: CEDEL2 – Aufsatz einer 61-jährigen Spanierin über ihre Heimat.³⁹¹

Im Korpus CEDEL2 finden sich Texte unterschiedlicher Länge, die vollständig angezeigt und sogar heruntergeladen werden können, doch die Anzahl der Lexeme wird erst nach dem Öffnen der Datei ersichtlich, wodurch eine Auswahl der Texte nach der passenden Länge erschwert wird. Leider enthält der bisher im Internet verfügbare Teil des Lernerkorpus keine Dateien mit Fehleranalyse. Es enthält lediglich Aufsätze von Lernern, die allerdings so abgespeichert sind, wie sie verfasst wurden, und um die oben genannten Angaben zur Person und zum jeweiligen Text erweitert wurden.

³⁹¹ Lozano, Cristóbal: *CEDEL2. Es_wr61_1_mda*. Online Publikation: http://cedel2.lernercorpora.com/student/es_wr_61_1_mda/?search_term=paisaje (16.04.2018).

2.3. Sprachenauswahl

Dieses Kriterium unterscheidet zwischen monolingualen und bi- beziehungsweise multilingualen Korpora. Viele Korpora des Spanischen sind monolingual, beziehen sich also nur auf die spanische Sprache, wie beispielsweise die Korpora der *Real Academia Española*, also das CREA (siehe Kapitel 2.1.1.1.) und das CORPES (siehe Kapitel 2.1.1.2.). Auch diachrone Korpora, die verschiedene Entwicklungsstufen derselben Sprache dokumentieren, wie das CORDE (siehe Kapitel 2.2.1.1.), bezeichnet man für gewöhnlich als monolingual. Die beiden Korpora des CdE (siehe Kapitel 2.1.1.3., 2.2.1.3. und 2.6.1.) sind ebenso monolingual. Im Gegensatz dazu gibt es die bi- oder multilingualen Korpora, welche auch Übersetzungskorpora genannt werden.

Tognini-Bonelli beschreibt Übersetzungskorpora als Sammlungen von Texten, die Übersetzungen eines Originaltexts darstellen, welcher selbst enthalten sein kann oder auch nicht. Ein Übersetzungskorpus als multilinguale Quelle kann den Übersetzungsprozess genauer aufzeigen. Am häufigsten wird diese Art von Korpus allerdings verwendet, um verschiedene Übersetzungen miteinander zu vergleichen sowie Gemeinsamkeiten und Unterschiede festzustellen, welche in einem einzelsprachlichen Korpus nicht entdeckt werden können. Es wird also meistens die Übersetzung als Produkt und nicht als Prozess untersucht. Innerhalb der Übersetzungskorpora wird zwischen Parallelkorpora, Vergleichskorpora und Korpora mit freien Übersetzungen unterschieden. Die ursprünglichen Parallelkorpora waren in der Tat parallel, das bedeutet, dass sich die Übersetzungen der Texte in solchen Korpora Satz für Satz entsprachen.³⁹² Heute sind Parallelkorpora meist auf Satz- oder Absatzebene aligniert und werden überwiegend für kontrastive linguistische Studien verwendet.³⁹³ Ein Parallelkorpus kann allerdings nicht nur Texte in verschiedenen Sprachen beinhalten, sondern auch unterschiedliche Varietäten einer Sprache.³⁹⁴

Teubert/Čermáková beschreiben Übersetzungskorpora, welche sie mit Parallelkorpora gleichsetzen, als besonders vorteilhaft, um herauszufinden, worin sich beispielsweise das Englische in Übersetzungen von realem Englisch unterscheidet. Diese Korpora geben auch Aufschluss darüber, wie bestimmte Lexeme einer Sprache, die in

³⁹² Vgl. Tognini-Bonelli (2001): op. cit., S. 6.

³⁹³ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 189.

³⁹⁴ Vgl. Tognini-Bonelli (2001): op. cit., S. 7-8.

2. Kategorisierung der Korpora

einer anderen Sprache zwei beinahe synonyme Entsprechungen haben, je nach ihrem Kontext korrekt übersetzt werden, wie beispielsweise das französische *travail* als englisches *work* oder *labour*.³⁹⁵

Aufgrund der Zweifel an der Repräsentativität strikter Parallelkorpora und der geringen Auswahlmöglichkeiten bezüglich des auf rigorose Art übersetzten Materials kam es zur Bildung der Vergleichskorpora. Bei Vergleichskorpora werden die jeweiligen Komponenten ausgewählt, weil sie sich bezüglich der externen Kriterien, wie beispielsweise der Einteilung in gesprochene oder geschriebene Sprache oder in verschiedene Register, auf vergleichbare Diskursbereiche beziehen. Hier entsprechen sich die Texte zwar nicht Satz für Satz und können daher nicht aligniert werden, aber es können Vergleiche zwischen den wichtigsten linguistischen Eigenschaften durchgeführt werden.³⁹⁶ Lemnitzer/Zinsmeister verweisen allerdings darauf, dass man unter dem Begriff Vergleichskorpus in der Übersetzungswissenschaft auch ein monolinguales Korpus versteht, das sowohl Originaltexte in der jeweiligen Sprache enthält als auch in diese Sprache übersetzte Texte. Erstere dienen dem Korpus als Hintergrund, vor dem Besonderheiten von Übersetzungen beobachtet werden können.³⁹⁷ Auch Lernerkorpora bestehend aus Originaltexten und den Übersetzungen der Schüler in der Zielsprache können als Parallel- oder Vergleichskorpora angesehen werden und für den Unterricht sehr nützlich sein.

Zwischen den Parallelkorpora (die aus wortwörtlichen Übersetzungen bestehen und deren Texte in einer Übersetzungsbeziehung zueinander stehen) und den Vergleichskorpora (in denen sich die Texte nicht Wort für Wort entsprechen) gibt es die Korpora der freien Übersetzung, welche Einsichten in eine kohärentere Art der Übersetzung bieten. Diese Korpora zeichnen sich durch die gegenseitige Übersetzungsbeziehung aus, entsprechen sich aber nicht Wort für Wort.³⁹⁸

Generell gibt es leider sehr wenige große Übersetzungskorpora der spanischen Sprache. Ein Problem der Parallelkorpora stellt beispielsweise die genaue Zuordnung dar: Um mit einem Parallelkorpus angemessen arbeiten zu können, muss dieses erst aligniert werden, sodass eine Einheit in der einen Sprache derjenigen in der anderen Sprache

³⁹⁵ Vgl. Teubert/ Čermáková (2007): op. cit., S. 73-74.

³⁹⁶ Vgl. Tognini-Bonelli (2001): op. cit., S. 7.

³⁹⁷ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 104.

³⁹⁸ Vgl. Tognini-Bonelli (2001): op. cit., S. 7.

entspricht. Meist wird hierfür, wie bereits erwähnt, die Satzgrenze verwendet. Diese ist aber nicht immer einfach zu bestimmen, da Abkürzungen, Ordinalzahlen und unterschiedliche satzaufspaltende und –vereinende Übersetzungsmethoden die automatische Ausrichtung erschweren und eine zeitaufwändige manuelle Einteilung unabdingbar machen. Die Ausrichtung auf lexikalischer Ebene ist noch problematischer. Aus eben diesem Grund existieren bisher nur wenige Parallelkorpora in einer beträchtlichen Größe (mehr als 5 Mio. Lexeme je Sprache).³⁹⁹

Für das Spanische werden im Folgenden das dreisprachige Parallelkorpus *Corpus Trilingüe Paralelo Grial* vorgestellt sowie die hauptsächlich auf das Galicische bezogene Sammlung von Parallelkorpora CLUVI der *Universidade de Vigo*.

2.3.1. Corpus Trilingüe Paralelo Grial

Für das Spanische gibt es also ein dreisprachiges Parallelkorpus namens *Grial*, das aus knapp 2,3 Mio. Lexemen aus den Sprachen Englisch, Spanisch und Katalanisch besteht. Das Korpus wurde von der *Grup de Recerca Interuniversitari en Aplicacions Lingüístiques* (GRIAL) entwickelt, die eine Art Arbeitskreis der vier Universitäten Kataloniens darstellt (*Universitat Autònoma de Barcelona*, *Universitat de Barcelona*, *Universitat de Lleida* und *Universitat Oberta de Catalunya*). Das Ziel dieser interuniversitären Forschungsgruppe ist die Erstellung linguistischer Ressourcen für die angewandte Linguistik, vor allem im Bereich der Lexikographie, der Übersetzung und der Sprachverarbeitung. Ihre Forschung beschäftigt sich vor allem mit den Sprachen Spanisch und Katalanisch.⁴⁰⁰

Das *Corpus Trilingüe Paralelo Grial* wurde automatisch morphosyntaktisch annotiert und soll vor allem der Lehre der englischen Sprache in technischen Ausbildungsrichtungen zugutekommen.⁴⁰¹ Mithilfe des Korpus sollen einerseits die Lehrkräfte Texte und Beispiele für den Unterricht zur Verfügung haben, andererseits soll es den Schülern möglich sein, bestimmte Ausdrücke und grammatikalische Mittel des Englischen und ihrer jeweiligen Muttersprache – sei es Spanisch oder Katalanisch –

³⁹⁹ Vgl. Teubert/ Čermáková (2007): op. cit., S. 74-76.

⁴⁰⁰ Vgl. Grup de Recerca Interuniversitari en Aplicacions Lingüístiques (ed.): *Inicio*. Online Publikation: <http://grial.uab.es/index.php?idioma=es> (28.10.2016).

⁴⁰¹ Vgl. Grup de Recerca Interuniversitari en Aplicacions Lingüístiques (ed.): *Recursos. Corpus*. Online Publikation: <http://grial.uab.es/recursos.php?id=1&idioma=es> (21.12.2016).

2. Kategorisierung der Korpora

gegenüberstellen zu können.⁴⁰² Es ist im Internet kostenlos und ohne jegliche Anmeldung zugänglich (<http://grial.uab.es/sensem/cpg>).

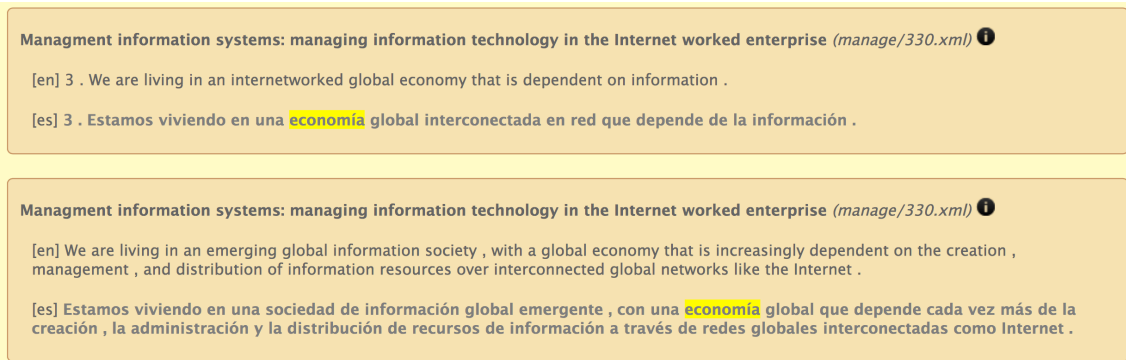


Abbildung 34: *Corpus Trilingüe Paralelo Grial* – Ergebnisse der Suche nach *economía* in einem Dokument über Management-Informationssysteme.⁴⁰³

Im *Corpus Trilingüe Paralelo Grial* werden die Abschnitte der spanischen und englischen Texte, in denen der Suchbegriff vorkommt und welche sich aufeinander beziehen, parallel dargestellt, um einen direkten Vergleich der Passagen zu ermöglichen.

2.3.2. CLUVI – *Corpus Lingüístico da Universidade de Vigo*

Außerdem gibt es das *Corpus Lingüístico da Universidade de Vigo* (CLUVI), welches im Internet frei und kostenlos zugänglich ist (<http://sli.uvigo.es/CLUVI/>). Diese Sammlung von Parallelkorpora der *Universidade de Vigo* umfasst ungefähr 25 Mio. Lexeme. In den verschiedenen Subkorpora des SLI (*Seminario de Lingüística Informática da Universidade de Vigo*) gibt es vor allem folgende Sprachkombinationen: Galicisch-Spanisch, Englisch-Galicisch-Französisch-Spanisch, Englisch-Galicisch, Französisch-Galicisch, Spanisch-Galicisch-Katalanisch-Baskisch, Baskisch-Spanisch und Portugiesisch-Englisch.⁴⁰⁴ Es handelt sich bei diesem im September 2003 online veröffentlichten Korpus also um eine Sammlung von Parallelkorpora, die sich mit dem Verhältnis des Galicischen zu verschiedenen europäischen Sprachen, aber auch zu den

⁴⁰² Vgl. Grup de Recerca Interuniversitari en Aplicacions Lingüístiques (ed.): *Proyectos. Corpus Trilingüe Paralelo GRIAL*. Online Publikation: <http://grial.uab.es/fproj.php?id=3&idioma=es> (21.12.2016).

⁴⁰³ Grup de Recerca Interuniversitari en Aplicacions Lingüístiques (ed.): *Corpus Trilingüe Paralelo GRIAL. Buscar 'economía' en 'manage/330'*. Online Publikation: <http://grial.uab.es/sensem/cpg/search?page=0&type=norm&text=econom%C3%ADa&idi=es&docs%5B%5D=8> (10.03.2017).

⁴⁰⁴ Vgl. Seminario de Lingüística Informática (SLI) da Universidade de Vigo (ed.): *Corpus Lingüístico da Universidade de Vigo: Pescudas no Corpus Paralelo CLUVI*. Online Publikation: <http://sli.uvigo.es/CLUVI/> (03.08.2011).

weiteren regionalen Amtssprachen Spaniens befasst. Das Hauptaugenmerk liegt dabei auf dem gesprochenen und geschriebenen Galicisch.⁴⁰⁵

Diese offene Sammlung paralleler Textkorpora wurde unter der Leitung von Xavier Gómez Guinovart entwickelt. Die sechs größten Parallelkorpora beziehen sich auf fünf spezielle Bereiche: Fiktion, EDV, Populärwissenschaft, Rechts- und Verwaltungswesen.⁴⁰⁶ Auf der Internetseite des CLUVI stehen einige Subkorpora mit verschiedenen Sprachkombinationen zur Verfügung, unter denen vor allem das CONSUMER-Korpus von Interesse ist für die vorliegende Arbeit. Hierbei handelt es sich um ein mehr als 5,5 Mio. Lexeme umfassendes spanisch-galicisch-katalanisch-baskisches Korpus der Verbraucherinformationen.⁴⁰⁷ Seit kurzem sind weitere kleine Parallelkorpora verfügbar, wie beispielsweise das DEGA-Korpus, in dem sogar die deutsche Sprache anhand literarischer Texte mit etwas mehr als 366.000 Lexemen mit dem Galicischen verglichen werden kann.⁴⁰⁸ Die Anzahl der alignierten Texte und Sprachpaare wächst ständig, da das CLUVI ein akademisches Forschungsprojekt ist, an dem weiterhin gearbeitet wird.⁴⁰⁹

In den Korpora des CLUVI können sowohl einfache als auch komplexe Suchen nach einzelnen Lexemen oder Phrasen durchgeführt werden. Die mehrsprachigen Entsprechungen der Suchausdrücke werden im Kontext der Übersetzungen dargestellt. Dabei kann nach Ausdrücken in einer beliebigen, im jeweiligen Korpus verfügbaren Sprache gesucht werden.⁴¹⁰

Bei einer Suche nach *pescado* im spanisch-galicisch-katalanisch-baskischen CONSUMER-Korpus finden sich unter anderem folgende Beispiele:

⁴⁰⁵ Vgl. Seminario de Lingüística Informática (SLI) da Universidade de Vigo (ed.): *Corpus Lingüístico da Universidade de Vigo: Pescudas no Corpus Paralelo CLUVI*. Op. cit. (03.08.2011).

⁴⁰⁶ Vgl. Universidade de Vigo. Grupo de Investigación Talg (ed.): *CLUVI Parallel Corpus*. Online Publikation: http://lod.iula.upf.edu/resources/metadata_CorpusCLUVIv2 (22.11.2016).

⁴⁰⁷ Vgl. Seminario de Lingüística Informática (SLI) da Universidade de Vigo (ed.): *Corpus CLUVI – Información sobre o Corpus CLUVI*. Online Publikation: <http://sli.uvigo.es/CLUVI/index.php> (20.08.2018).

⁴⁰⁸ Vgl. Seminario de Lingüística Informática (SLI) da Universidade de Vigo (ed.): *Search the CLUVI Parallel Corpus. Corpus CLUVI sections*. Online Publikation: http://sli.uvigo.es/CLUVI/index_en.html (23.01.2017).

⁴⁰⁹ Vgl. Seminario de Lingüística Informática (SLI) da Universidade de Vigo (ed.): *CLUVI Parallel Corpus – More Información. Introduction*. Online Publikation: http://sli.uvigo.es/CLUVI/info_en.html (23.01.2017).

⁴¹⁰ Vgl. Seminario de Lingüística Informática (SLI) da Universidade de Vigo (ed.): *CLUVI Parallel Corpus – More Información. Introduction*. Op. cit. (23.01.2017).

2. Kategorisierung der Korpora

18- CM1 (4895)	
ES	La carne y el pescado deben aparecer con la misma frecuencia, y, mientras no se sufran problemas de colesterol, se pueden tomar seis huevos a lo largo de toda la semana.
GL	A carne e o peixe deben aparecer coa mesma frecuencia, e, mentres non se sufran problemas de colesterol, pódense tomar seis ovos ó longo de toda a semana.
CA	La carn i el peix han d'aparèixer amb la mateixa freqüència i, mentre no es pateixin problemes de colesterol, es poden prendre sis ous al llarg de tota la setmana.
EU	Haragia eta arraina berdintsu jan behar dira eta kolesterol arazorik izan ezik, sei arrautza jan daitezke astean zehar.

Abbildung 35: CLUVI – Ergebnis des CONSUMER-Korpus bei der Suche nach *pescado*.⁴¹¹

Während in der ursprünglichen Version des CLUVI die Textausschnitte der einzelnen Sprachen beziehungsweise Varietäten parallel nebeneinander angeordnet wurden, werden sie in der neuen Version untereinander aufgelistet. Diese Darstellung ist zunächst gewöhnungsbedürftig, da nicht mehr auf einen Blick gesehen werden kann, welche Texte in allen vier Sprachen verfügbar sind. Nun muss der Nutzer die Liste der Textbeispiele von oben nach unten durchsehen, um passende Texte zu finden. Ein Vorteil der neuen Version ist jedoch, dass der Suchbegriff farblich hervorgehoben ist und die Entsprechungen desselben in den anderen Sprachen ebenso farblich markiert werden. Dies erleichtert insbesondere das Auffinden der baskischen Pendants, welche nur schwer ausfindig gemacht werden können, sofern der Nutzer über keinerlei Kenntnis dieser Sprache verfügt.

2.4. Medium

Zusätzlich unterscheidet man in geschriebene, gesprochene und multimodale Korpora, wie z. B. Videokorpora. Es handelt sich hierbei um das Medium, in dem die Primärdaten entstanden sind. Lemnitzer/Zinsmeister geben an, dass zwar bei Korpora gesprochener Sprache zwischen den Sprachsignalen und den Transkripten unterschieden werden muss, sie zählen Letztere jedoch ebenfalls zu den Korpora gesprochener Sprache. Die Unterscheidung zwischen gesprochener und geschriebener Sprache ist nicht immer leicht zu treffen, wie man an Hörspieltexten, Theaterstücken oder Texten der Chatsprache erkennen kann.⁴¹² Hierbei muss laut Koch/Österreicher generell zwischen medialer und

⁴¹¹ Seminario de Lingüística Informática (SLI) da Universidade de Vigo (ed.): *CLUVI Parallel Corpus – Buscar ‚pescado‘ en el corpus CONSUMER*. Online Publikation: <http://sli.uvigo.gal/CLUVI/index.php?corpus=20&tipo=14&lang=en> (06.03.2018).

⁴¹² Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 104.

konzeptioneller Mündlichkeit beziehungsweise Schriftlichkeit unterschieden werden. Bei der Chatsprache oder bei Sprechblasen in Comics ist die Äußerung zwar graphisch realisiert (schriftliches Medium), sie entspricht jedoch nicht unserer Vorstellung von Schriftlichkeit (mündliche Konzeption).⁴¹³ Für Lemnitzer/Zinsmeister dient in Bezug auf die Arbeit mit Korpora vor allem die mediale Schriftlichkeit und Mündlichkeit als Unterscheidungskriterium.⁴¹⁴

2.4.1. Korpora geschriebener Sprache

Generell sind die heute verfügbaren Korpora meist solche der geschriebenen Sprache, da schriftliche Dokumente leichter verfügbar sind. Von den bereits beschriebenen Korpora beinhalten beispielsweise die historischen Korpora (siehe Kapitel 2.2.1.) ausschließlich geschriebene Texte und auch Parallelkorpora, wie das CLUVI der *Universidade de Vigo* (siehe Kapitel 2.3.2.), beschränken sich darauf.

Umfassende Korpora wie das CREA (zumindest in der Version von 2015, siehe Kapitel 2.1.1.1.) oder auch das CdE:Orig (siehe Kapitel 2.1.1.3. und 2.2.1.3.) enthalten sowohl geschriebene als auch gesprochene Texte, wobei letztere im Korpus immer als geschriebene Texte dargestellt werden. Unter den Korpora des Spanischen stellt das CORPES (siehe Kapitel 2.1.1.2.) in Bezug auf das Medium eine Besonderheit dar, da es geschriebene und gesprochene Texte beinhaltet und letztere nicht nur als Transkriptionen zur Verfügung stehen, sondern auch als Tonspur. Man kann sich somit, wie im Abschnitt über das CORPES zu sehen war, die Konkordanzen einer Suchanfrage anzeigen lassen und den jeweiligen Ausschnitt anhören, sofern es sich um einen der gesprochenen Texte handelt.

Obgleich der Großteil der verfügbaren Korpora sich mit geschriebener Sprache befasst und die Korpora der gesprochenen Sprache somit die Besonderheit darstellen, sollen im Folgenden zunächst zwei Korpora der geschriebenen Sprache vorgestellt werden – das *Corpus del Español Actual* und das *Corpus periodístico Hemero* – bevor ausführlich auf mündliche und multimodale Korpora eingegangen wird.

⁴¹³ Vgl. Koch/ Oesterreicher (2011): op. cit., S. 3.

⁴¹⁴ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 104.

2.4.1.1. CEA – *Corpus del Español Actual*

Das *Corpus del Español Actual* (kurz: CEA) ist ein umfangreiches Korpus des geschriebenen, zeitgenössischen Spanisch aus Akten, Gesetzestexten und Wikipedia-Einträgen. Das 540 Mio. Lexeme umfassende opportunistische Korpus wurde von Carlos Subirats und Marc Ortega von der *Universidad Autónoma de Barcelona* im Jahr 2012 erstellt und setzt sich aus drei Subkorpora zusammen: dem spanischen Teil des *Europarl: European Parliament Proceedings Parallel Corpus* (1996-2010), dem spanischen Teil des *Wikicorpus* (2006) und dem spanischen Teil des *MultiUN: Multilingual UN Parallel Text* (2000-2009). Es enthält aktuelle Texte im HTML-Format aus den Jahren 1996 bis 2010, welche mittels der Suchmaske *CQPweb* durchsucht werden können. Das lemmatisierte sowie morphosyntaktisch annotierte Korpus ermöglicht Suchen nach Lexemen, Lemmata, Wortarten und Konstruktionen unter Verwendung regulärer Ausdrücke. Zudem können Frequenzlisten und Kollokationen abgerufen werden. Es ist kostenlos und frei im Internet zugänglich (<http://spanishfn.org/tools/cea/spanish>).⁴¹⁵

Anhand des CEA kann mittels unterschiedlicher Verfahren gesucht werden. Es können Standardsuchen nach Lexemen oder Phrasen durchgeführt werden, wobei zwischen Groß- und Kleinschreibung unterschieden und entweder das komplette Korpus oder ein zuvor genauer bestimmtes Subkorpus durchsucht werden kann. Zudem ist die Nutzung von *Wildcards*, Metazeichen, Wortartentags und regulären Ausdrücken möglich.

Eine besondere Option ist *Word lookup*. In dieser Funktion kann nach einer Wortform gesucht werden unter der Angabe, ob die Treffer mit dieser Form beginnen, damit enden, diese beinhalten oder genau diese Wortform repräsentieren sollen. Wenn der Nutzer nun beispielsweise nach der Form *trabajo* sucht, unter der Prämisse, dass die Treffer genau diese Form repräsentieren sollen, so werden drei Treffer angezeigt: *trabajo* als Verb im Indikativ, als Substantiv und als Eigenname. Außerdem werden die Häufigkeit und der Prozentsatz, der auf die jeweilige Möglichkeit entfällt, angegeben:

⁴¹⁵ Vgl. Subirats, Carlos/ Ortega, Marc: *Corpus del Español Actual (CEA)*. Online Publikation: <http://spanishfn.org/tools/cea/spanish> (17.03.2017).

No.	Search result
1	trabajo_VPRED_IND
2	Trabajo_NPROP
3	trabajo_N

Abbildung 36: CEA – Ergebnisse der Suche nach *trabajo* in der Funktion *Word Lookup*.⁴¹⁶

No. of occurrences	Percent
536562	60.97%
290458	33%
53065	6.03%

Abbildung 37: CEA – Häufigkeit der Ergebnisse der Suche nach *trabajo* in der Funktion *Word Lookup*.⁴¹⁷

Bei der Anzeige der Ergebnisse muss jedoch hinzugefügt werden, dass die hohe Anzahl der Fälle mit *Trabajo* als Eigenname der Beschaffenheit des Korpus aus Akten und Gesetzestexte geschuldet ist. Diese Funktion ist sehr nützlich, um ambigüe Formen zu disambiguieren und sich verschiedene Verwendungsweisen anzeigen zu lassen.

Außerdem kann die Frequenzliste des CEA nach Wortformen, Lemmata, Wortarten oder Wortbeginn geordnet angezeigt werden und es kann speziell nach Lexemen mit bestimmten Häufigkeiten gesucht werden. Derartige Listen können entweder in absteigender oder aufsteigender Reihenfolge, der Häufigkeit nach oder alphabetisch angezeigt werden. Zudem können auch eigene Frequenzlisten erstellt werden.

2.4.1.2. *Corpus periodístico Hemero*

Ein weiteres opportunistisches Korpus ist das *Corpus periodístico Hemero* (kurz: *Hemero*), welches von *Molino de Ideas*, einem Unternehmen für natürliche Sprachverarbeitung, erstellt wurde.⁴¹⁸ Es handelt sich hierbei um ein schriftliches Korpus der spanischen Standardsprache, welches verschiedenste Texte spanischsprachiger

⁴¹⁶ Subirats, Carlos/ Ortega, Marc: *Corpus del Español Actual (CEA). Word lookup – Buscar ,trabajo'*. Online Publikation: <http://sfncorpora.uab.es/CQPweb/cea/redirect.php?redirect=concBreakdownBoth&qname=f8b81grfyp&p=50&uT=y> (17.03.2017).

⁴¹⁷ Subirats/ Ortega: *Corpus del Español Actual (CEA). Word lookup – Buscar ,trabajo'*. Op. cit. (17.03.2017).

⁴¹⁸ Vgl. Molino de Ideas (ed.): *Molino de Ideas. Tecnología lingüística y Procesamiento de Lenguaje Natural*. Online Publikation: <http://molinodeideas.es/index.html> (14.03.2017).

2. Kategorisierung der Korpora

Zeitungen in sich vereint. Die Zeitungstexte stammen aus Spanien, Argentinien und Mexiko aus den Jahren 1997 bis 2009.⁴¹⁹

Im Internetauftritt des Korpus wird es als ein erstes Experiment der Bildung eines Zeitungskorpus beschrieben, welches derzeit mehr als 660 Mio. Lexeme umfasst und weiterhin wächst.⁴²⁰ Das Korpus ist kostenlos und ohne Anmeldung im Internet nutzbar (www.hemero.es).

Anhand der Benutzeroberfläche kann zunächst ein Subkorpus ausgewählt werden: *Periódicos de España*, *Periódicos de Argentina* oder *Periódicos de México*. Im zweiten Schritt wird bestimmt, ob in den Texten selbst oder den Titeln gesucht werden soll. Im Korpus *Hemero* kann entweder nach einem bloßen Lexem oder Satz, mit *Wildcard* (jedoch nicht zu Beginn des Lexems), nach ‚ähnlichen‘ Lexemen (wobei sogar der Grad der Ähnlichkeit angegeben werden kann) oder nach Ausdrücken in einem gewissen Abstand zum Suchbegriff gesucht werden.⁴²¹

Sofern beispielsweise nach dem Lexem *naturaleza* gesucht wird, findet das Korpus 10829 passende Dokumente und es werden nur Treffer angezeigt, die wortwörtlich den Begriff *naturaleza* beinhalten.⁴²² Wenn nun nach ‚sehr ähnlichen‘ Lexemen gesucht wird, findet die Suchmaschine 10916 passende Dokumente, wobei unter anderem Treffer wie *naturalezas* und **naturalaza* angezeigt werden.⁴²³ Sofern ‚sehr wenig ähnliche‘ Lexeme ermittelt werden sollen, zeigt das Korpus 156767 Dokumente an, denn es werden auch Lexeme wie *natural*, *naturalistas* oder *naturalia* als Treffer gewertet.⁴²⁴

⁴¹⁹ Vgl. Calvo, José (2016): *Textdaten Romanistik. Corpus periodístico Hemero*. Online Publikation: <http://wikis.fu-berlin.de/pages/viewpage.action?pageId=736854944> (14.03.2017).

⁴²⁰ Vgl. Molino de Ideas (ed.): *Hemero.es - El corpus molinero*. Online Publikation: <http://www.hemero.es> (12.07.2019).

⁴²¹ Vgl. Molino de Ideas (ed.): *Hemero.es - El corpus molinero*. Op. cit. (14.03.2017).

⁴²² Vgl. Molino de Ideas (ed.): *Hemero.es - El corpus molinero – Buscar ‚Una palabra o una frase‘ – naturaleza*. Online Publikation: [http://www.hemero.es/#1&espania\\$&texto&naturaleza](http://www.hemero.es/#1&espania$&texto&naturaleza) (14.03.2017).

⁴²³ Vgl. Molino de Ideas (ed.): *Hemero.es - El corpus molinero – Buscar ‚Una palabra y sus palabras parecidas‘ – ‚muy parecido‘ – naturaleza*. Online Publikation: [http://www.hemero.es/#3&espania\\$&texto&naturaleza&0.80](http://www.hemero.es/#3&espania$&texto&naturaleza&0.80) (14.03.2017).

⁴²⁴ Vgl. Molino de Ideas (ed.): *Hemero.es - El corpus molinero – Buscar ‚Una palabra y sus palabras parecidas‘ – ‚muy poco parecido‘ – naturaleza*. Online Publikation: [http://www.hemero.es/#3&espania\\$&texto&naturaleza&0.20](http://www.hemero.es/#3&espania$&texto&naturaleza&0.20) (14.03.2017).

Fecha	Resumen
2007	conservación de la naturaleza en el ámbito de la provincia de Zamora: naturalistas , biólogos y licenciados en
2007	conservación de la naturaleza en el ámbito de la provincia: naturalistas , biólogos y licenciados en Ciencias
2007	investigación científica del medio natural , la divulgación y difusión de los valores naturales , la educación
2009	investigación científica del medio natural , la divulgación y difusión de los valores naturales , la educación
2009	investigación científica del medio natural , la divulgación y difusión de los valores naturales , la educación
2007	obtención y consumo de alimentos que se hayan obtenido con prácticas respetuosas con la naturaleza ; y
2007	ritmos de la naturaleza . Para contactar C/ Mariquince, 8. naturalia_z@terra.es

Abbildung 38: *Hemero* – Suche nach *naturaleza* und ‚sehr wenig ähnlichen Lexemen‘.⁴²⁵

Die Suche nach ‚ähnlichen‘ Lexemen erscheint vor allem dann sinnvoll, wenn der Nutzer Artikel zu einem bestimmten Thema sucht, dieses Thema aber nicht auf ein Lexem eingrenzen möchte oder kann. Auf diese Weise werden verwandte Lexeme ebenfalls angezeigt, und durch die Möglichkeit der Angabe der Ähnlichkeit zum Suchbegriff kann die Suche erweitert oder genauer eingegrenzt werden.

Neben der Suche nach ‚ähnlichen‘ Lexemen ist eine weitere Besonderheit dieses Korpus die Verfügbarkeit der vollständigen Texte. Am linken Rand der Konkordanzen findet der Nutzer das Erscheinungsjahr des jeweiligen Beispiels und am rechten Rand kann er sich den vollständigen Text inklusive Quelle anzeigen lassen oder einem Link zum ursprünglichen Zeitungsartikel folgen.

2.4.2. Korpora gesprochener Sprache

Im Gegensatz zu den eben genannten Korpora befassen sich andere, wie das C-ORAL-ROM, ausschließlich mit gesprochenen Daten und sind damit eindeutig mündliche Korpora. Bei Korpora gesprochener Sprache müssen die Sprachdaten zunächst vom phonischen in das grafische Medium überführt werden. Diese ‚Transkription‘ kann mithilfe spezifischer Programme wie dem französischen

⁴²⁵ Molino de Ideas (ed.): *Hemero.es - El corpus molinero – Buscar ‚Una palabra y sus palabras parecidas‘ – ‚muy poco parecido‘ – naturaleza*. Op. cit. (14.03.2017).

2. Kategorisierung der Korpora

*TranscriberAG*⁴²⁶, *f4transkript*⁴²⁷ von der Universität Marburg oder dem *EXMARaLDA-Editor*⁴²⁸ von der Universität Hamburg durchgeführt werden. Auch das Programm *PRAAT*⁴²⁹, welches ursprünglich für phonetisch-prosodische Analysen entwickelt wurde, kann für Transkriptionen genutzt werden.⁴³⁰

Die bereits erwähnten Referenzkorpora CREA, CORPES und CdE:Orig (siehe Kapitel 2.1.1.) umfassen ebenfalls jeweils ein mündliches Subkorpus, um die spanische Sprache so gut wie möglich abbilden zu können. Leider ist das CORPES das einzige der drei Referenzkorpora, in dem auch Textausschnitte angehört werden können.

Stellvertretend für diese drei Korpora soll die Aufteilung des CREA im Folgenden genauer durchleuchtet werden, um den Aufbau des mündlichen Teils der Referenzkorpora zu erkennen: Das Korpus *CREA Oral*, das, wie bereits erwähnt, 10% des gesamten CREA ausmacht, umfasst knapp 9 Mio. Formen in mehr als 1600 Dokumenten. Die Texte sind zur Hälfte spanischer und zur Hälfte lateinamerikanischer Herkunft. Sie sind verschiedenen Ursprungs, entweder aus dem Radio oder Fernsehen, aus Direkterfassungen oder Telefonaten. Zudem wurden manche eigens erfasst oder stammen aus Texten, die zuvor transkribiert oder auch kodifiziert wurden. Die Texte aus Radio und Fernsehen lassen sich in Rubriken wie Nachrichten, Reportagen, Interviews, Sportsendungen oder Gewinnspiele einteilen. Diese Dateien aus Radio und Fernsehen stellen den Kern des mündlichen Teils dar, aber es finden sich auch andere Texte wie beispielsweise politische Diskurse, Telefongespräche, Mitteilungen auf Anrufbeantwortern oder informelle Dialoge. Bei den einzelnen Texten sind Informationen wie der Grad der Formalität, die Art des Zuhörers (Gesprächspartner oder passiver Zuhörer) und der Kanal (beispielsweise Gespräche sich gegenüberstehender Personen) vermerkt. In der Transkription finden sich auch Informationen bezüglich der Sprecher, des Sprecherwechsels und im Hinblick auf non-verbale Elemente.⁴³¹

⁴²⁶ Vgl. DGA (ed.2014): *TranscriberAG*. Online Publikation: <http://transag.sourceforge.net> (24.11.2016).

⁴²⁷ Vgl. Dr. Dresing & Pehl GmbH (ed.2016): *f4 & f5transkript*. Online Publikation: <http://www.audiotranskription.de/f4.htm> (24.11.2016).

⁴²⁸ Vgl. Hamburger Zentrum für Sprachkorpora (ed.): *EXMARaLDA*. Op. cit. (24.11.2016).

⁴²⁹ Vgl. Boersma, Paul/ Weenink, David: *Praat: doing phonetics by computer*. Online Publikation: <http://www.fon.hum.uva.nl/praat/> (24.11.2016).

⁴³⁰ Vgl. Kabatek/ Pusch (2011): op. cit., S. 211.

⁴³¹ Vgl. Real Academia Española (ed.): *Banco de datos. CREA Oral*. Online Publikation: <http://www.rae.es/rae/gestores/gespub000019.nsf/voTodosporId/F651954693D3AC5EC125716400426004?OpenDocument> (20.05.2011).

2.4.2.1. *Macrocorpus de la norma lingüística culta de las principales ciudades del mundo hispánico*

Das *Macrocorpus de la norma lingüística culta de las principales ciudades del mundo hispánico* (kurz: *Macrocorpus*) ist ein Korpus, das sich mit der gesprochenen Sprache beschäftigt. Das *Macrocorpus* bietet die Transliteration von 84 Stunden gesprochener spanischer Sprache. Es handelt sich dabei um 168 Sprecher gehobenen Niveaus aus zwölf spanischsprachigen Städten, darunter neun lateinamerikanische Metropolen (México, Caracas, Santiago de Chile, Santafé de Bogotá, Buenos Aires, Lima, San Juan de Puerto Rico, La Paz und San José de Costa Rica) sowie Madrid, Sevilla und Las Palmas de Gran Canaria. Es wurden Interviews von einer halben Stunde Dauer mit jeweils 14 Bewohnern jeder Stadt durchgeführt, bei welchen aus Gründen der Vergleichbarkeit immer auf ähnliche Weise vorgegangen wurde. Das *Macrocorpus* wurde im Jahr 1998 auf CD-ROM veröffentlicht und ist leider ausschließlich über dieses Medium erhältlich.⁴³²

BA-3. Mujer de 28 años. Profesora de Filosofía

Section Break (Continuous)

Inf.- ¿Te animás a escucharlo de nuevo?

Enc.- Sí, bueno; o si no, contáme... a ver... ¿Sonia qué hace o qué...? Andá ubicándome más o menos qué están haciendo todas [...].

Inf.- Bueno. Sonia terminó este año... eso creo.

Enc.- ¿Terminó la carrera?

Inf.- Sí, yo te había dicho.

Enc.- ¿Qué terminó?, no.

Inf.- ¿No?, sí... terminó... a ver... en julio... creo que terminó... y terminó... Dio el examen

Abbildung 39: *Macrocorpus* – Ausschnitt aus einem Text aus Buenos Aires.⁴³³

Die Autoren des *Macrocorpus* nennen im Begleittext zu ihrer CD-ROM einige Gründe, warum auch die Texte der Korpora gesprochener Sprache in gewöhnlicher Schreibung und nicht als phonetische Transkription gespeichert werden sollten. Erstens müsste man ein Experte sein, um die Texte in phonetischer Transkription entschlüsseln

⁴³² Vgl. Samper Padilla (1998): op. cit.

⁴³³ Samper Padilla (1998): op. cit.

und für eigene Studien verwenden zu können. Die jeweils angewandten Transkriptionsregeln müssten dem Nutzer zudem bekannt und genügend vertraut sein, um darin zu forschen. Zweitens ist die konventionelle Orthographie besser geeignet für Untersuchungen in den Bereichen der Morphologie, Syntax, Lexik sowie Diskurslinguistik, und sie kann zudem einfacher für spätere Statistiken in Bezug auf Lexik und Konkordanzen genutzt werden. Die phonetische Transkription ist außerdem sehr zeitaufwändig, weshalb sie bei umfassenden Korpora kaum praktikabel ist. Dennoch sei die Verwendung der phonetischen Transkription bei Korpora von begrenztem Umfang mit einer bestimmten phonetischen Zielsetzung verständlich, laut der Autoren. In diesem Fall scheine eine zusätzliche orthographische Transliteration jedoch hilfreich.⁴³⁴

2.4.2.2. C-ORAL-ROM

Dieses Korpus wird auf der zugehörigen Internetseite als integriertes Referenzkorpus für gesprochene romanische Sprachen bezeichnet.⁴³⁵ Es ist ein Vergleichskorpus und umfasst je 300.000 Lexeme spontaner, gesprochener Sprache des Französischen, Italienischen, Portugiesischen und Spanischen. Die Subkorpora sind nach demselben Muster erstellt worden, und die Texte wurden mit derselben Methode gesammelt. Letztere sind mit Informationen über die Herkunft, die Sprecher und eventuelle Sprecherwechsel versehen sowie prosodisch annotiert.⁴³⁶ Auf der CD-ROM sind die Tonsequenzen und die Transkriptionen erfasst, die einzelnen Subkorpora, in denen die Texte nach Formalitätsgraden und Kommunikationssituationen geordnet sind, haben allerdings einen geringen Umfang, was teilweise durch die rein sprechsprachliche Ausrichtung erklärt werden kann.⁴³⁷ Leider ist das Korpus ausschließlich auf CD-ROM verfügbar.⁴³⁸ Positiv zu vermerken ist allerdings, dass das C-ORAL-ROM mündliche Texte zu einer Vielzahl unterschiedlicher Themen in den verschiedensten Bereichen des alltäglichen Lebens sowohl in Form von formellen als auch informellen Texten abdeckt.

⁴³⁴ Vgl. Samper Padilla (1998): op. cit.

⁴³⁵ Vgl. Directorate General Information Society (ed.): *Integrated reference corpora for spoken romance languages: C-ORAL-ROM – Intro*. Online Publikation: <http://lablita.dit.unifi.it/coralrom/intro.html> (03.12.2011).

⁴³⁶ Vgl. John Benjamins Publishing Company (ed.): *C-ORAL-ROM. Integrated Reference Corpora for Spoken Romance Languages*. Online Publikation: <https://www.benjamins.com/#catalog/books/scl.15/main> (22.12.2016).

⁴³⁷ Vgl. Directorate General Information Society (ed.): op. cit. (03.12.2011).

⁴³⁸ Vgl. Cresti, Emanuela/ Moneglia, Massimo (Hrsg.) (2005): *C-ORAL-ROM. Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam [u.a.]: Benjamins.

Es finden sich beispielsweise Nachrichtensendungen, Wetterberichte, Sportreportagen und Talkshowausschnitte, aber auch politische Reden und Debatten, Vorlesungen und Konferenzen.⁴³⁹

Koch/Oesterreicher beschreiben das C-ORAL-ROM als einen bemerkenswerten Schritt nach vorn, da es Korpora des phonisch realisierten Französisch, Spanisch, Italienisch und Portugiesisch in doppelter digitaler Version beinhaltet, welche angehört und gelesen werden können.⁴⁴⁰

Im Internet kann man sich die jeweils ersten 30 Sekunden der im C-ORAL-ROM enthaltenen Tonaufnahmen anhören und die erste Seite des transkribierten Texts ansehen (<http://www.elda.org/en/proj/coral/demo/index.html>).⁴⁴¹ Sollte man also über die Anschaffung der CD-ROM nachdenken, sich aber nicht sicher sein, kann man zunächst in einige Dateien hineinhören und sich so einen ersten Eindruck von der Korpuserstellung und den im C-ORAL-ROM enthaltenen Texte machen.

⁴³⁹ Vgl. Cresti/ Moneglia (Hrsg.) (2005): op. cit.

⁴⁴⁰ Vgl. Koch/ Oesterreicher (2011): op. cit., S. 38.

⁴⁴¹ Vgl. C-ORAL-ROM Project (ed.): *Corpus Demo*. Online Publikation: <http://www.elda.org/en/proj/coral/demo/index.html> (22.12.2016).

@Title: dialogo sobre fútbol y baloncesto
@File: efamd103
@Participants: MIG, Miguel, (man, B, 2, taxi driver, x, Madrid)
PAC, Paco, (man, A, 2, air condition installer, x, Madrid)
@Date: 04/03/2001
@Place: Madrid
@Situation: conversation at home
@Topic: football and basketball
@Source: C-ORAL-ROM
@Class: informal, familiar/private, dialogue
@Length: 7' 08"
@Words: 1589
@Acoustic_quality: A
@Transcriber: Guillermo
@Revisor: Jesús; Manuel, Inma, Jesús, Ana and Guillermo (prosody)
@Comments:



*MIG: tampoco hay mucho que [/] no ? que escuchar // xxx //
*PAC: no sé / &eh / has visto el Marca hoy ?
*MIG: no // el Marca no lo he &com
*PAC: ponía [/] decía Gaspard / que le habían robado la cartera //
*MIG: < hhh // >
%act: (1) laugh
*PAC: [<] < por lo del gol // > ponía le &ha [/] le han [/] me han robado la cartera //
*MIG: y qué [/] y qué ha dicho él ?
*PAC: pues que no // que decían que era un gol /
*MIG: que es < legal // >
*PAC: [<] < / que era > válido // < hhh >
*MIG: [<] < ya // >
*PAC: / que lo ha anulado / porque se ha equivocado el árbitro / pero que era válido //

Abbildung 40: C-ORAL-ROM – Ausschnitt eines informellen, privaten Dialogs.⁴⁴²

Anhand dieses Ausschnitts kann man erkennen, dass in den Transkriptionen des C-ORAL-ROM zunächst einige Informationen zu Thema, Sprechern, Situation, Aufnahme und Bearbeitung aufgeführt werden, bevor der eigentliche Text beginnt. Durch einen Klick auf den Lautsprecher kann sich der Nutzer das Dokument gleichzeitig anhören.

⁴⁴² C-ORAL-ROM Project (ed.): *Corpus Demo. Diálogo informal, privado (efamd103)*. Online Publikation:
http://www.elda.org/en/proj/coral/demo/spain/Informal/Family_Private/dialogues_conversations/efamd103.html (10.03.2017).

2.4.2.3. CORLEC – *Corpus Oral de Referencia de la Lengua Española Contemporánea*

Das *Corpus Oral de Referencia de la Lengua Española Contemporánea* (CORLEC) ist ein mündliches Korpus, das 1,1 Mio. Lexeme aus Texten gesprochener Sprache enthält. Hierfür wurden Dokumente aus unterschiedlichen Bereichen, beispielsweise der Verwaltung, der Erziehung, der Politik oder der Wissenschaft, von einem Team transkribiert.⁴⁴³ Die transkribierten Texte sind im Internet frei verfügbar, und das Korpus kann zudem entweder in UTF8 oder ASCII heruntergeladen werden (<http://www.llf.uam.es/ESP/Corlec.html>).⁴⁴⁴

Für die Transkription der Dokumente wurden spezielle *tags* entwickelt (zum Begriff *tag* siehe Kapitel 1.4.5.), welche Besonderheiten der gesprochenen Sprache genauer beschreiben. So wurde beispielsweise angegeben, wenn Lexeme abgekürzt wurden, Sprecher zögerten, Bestätigung/ Zweifel/ Verneinung nonverbal ausdrückten, Geräusche von sich gaben oder lautmalerische Ausdrücke gebrauchten. Im Folgenden werden ein paar Beispiele für derartige Etikettierungen angegeben.⁴⁴⁵

Abkürzungen	<i>Está ahí al la<(d)>o, pa<(r)><(a)> allá, to<(d)><(o)>, etc.</i>
Geräusche	<risas>, <aplausos>, <música>
Onomatopoetika	<HI> Y los coches <onomatopéyico>piun, piun, piun, piun, piun</onomatopéyico>

Abbildung 41: CORLEC – Beispiele für Etikettierungen.⁴⁴⁶

Alle Dokumente des CORLEC sind im Internet frei zugänglich, doch leider gibt es mangels einer Suchmaske keine Möglichkeit, im Korpus zu suchen. Zudem sind die Audiodateien nicht verfügbar.

⁴⁴³ Vgl. Laboratorio de Lingüística Informática de la Universidad Autónoma de Madrid (ed.): *Corpus Oral de Referencia de la Lengua Española Contemporánea (CORLEC). Información y descripción del corpus*. Online Publikation: <http://www.llf.uam.es/ESP/Info%20Corlec.html> (24.11.2016).

⁴⁴⁴ Vgl. Laboratorio de Lingüística Informática de la Universidad Autónoma de Madrid (ed.): *Corpus Oral de Referencia de la Lengua Española Contemporánea (CORLEC)*. Online Publikation: <http://www.llf.uam.es/ESP/Corlec.html> (25.10.2016).

⁴⁴⁵ Vgl. Laboratorio de Lingüística Informática de la Universidad Autónoma de Madrid (ed.): *Corpus Oral de Referencia de la Lengua Española Contemporánea (CORLEC)*. Op. cit. (24.11.2016).

⁴⁴⁶ Vgl. Laboratorio de Lingüística Informática de la Universidad Autónoma de Madrid (ed.): *Corpus Oral de Referencia de la Lengua Española Contemporánea (CORLEC). Información y descripción del corpus*. Op. cit. (24.11.2016).

2. Kategorisierung der Korpora

<ACON021A.WPT> <También en cinta nº8>
<22-5-91>
<fuente=conversación telefónica>
<localización=Madrid>
<términos=plan, tenis, estudiantes extranjeros, trabajo, exámenes, agobio, fiesta, canciones, idioma alemán, lengua, libros, morfología>
<H1=Mujer, profesora de español para extranjeros en la Universidad, (filóloga), 23 años>
<H2=Mujer, filóloga, estudiante, 23 años>

<texto>

<H1> Digo: "Bueno". Ya se me ha estropeado el plan.
<H2> Jo.
<H1> ¡Me he podido quedar! Si me pod<palabra cortada>... Me hubiese podido quedar, pero tendría que haber vuelto, ¿no? y ya me ha parecido demasiado. Digo: "Bah, ya... ya jugaré otro día con él". Porque me... alirme, me dice Alberto, dice: "¿no te quedas a jugar un poco más?"
<H2> Pero, ¿ése Alberto es... es un alumno tuyo... extranjero?
<H1> Sí.
<H2> ¿Y de dónde es, llamándose Alberto?
<H1> Italiano.
<H3> Ah, italiano. Claro, claro.
<H1> Di<palabra cortada>... tiene, fíjate, se llama Alberto y tiene un apellido catalán.
<H2> <risas> ¿Y eso?
<H1> ¿Eh?
<H2> ¿Y eso?
<H1> Pues no sé. Digo: "¡Pero bueno!" Dice: "Sí, sí", dice: "Si... muchas veces me he hecho pasar por español"
<H2> <risas>

Abbildung 42: CORLEC – Beispieltext zur Veranschaulichung der Transkription.⁴⁴⁷

Auch in diesem Korpus werden zunächst einige Daten zu Sprechsituation, Ort, Schlagwörtern und Sprechern angegeben, bevor die Transkription des Texts angezeigt wird.

2.4.2.4. C-Or-DiAL – *Corpus Oral Didáctico Anotado Lingüísticamente*

Das *Corpus Oral Didáctico Anotado Lingüísticamente* (C-Or-DiAL) des *Laboratorio de Lingüística Italiana* (Lablita) der *Università degli Studi Firenze* ist ein für die Lehre und Forschung an der spanischen Sprache entwickeltes Korpus der gesprochenen Standardsprache. Es handelt sich dabei um keine künstlichen, sondern reale Aufnahmen, welche die linguistische Wirklichkeit des Spanischen wiedergeben sollen.⁴⁴⁸

Das C-Or-DiAL besteht aus 240 Aufnahmen mit insgesamt knapp 10 Stunden Sprachmaterial inklusive dazugehöriger Transkripte, welche sich auf knapp 119.000 Lexeme belaufen. Die Mehrzahl der Aufnahmen wurde ab dem Jahr 2004 in Madrid

⁴⁴⁷ Laboratorio de Lingüística Informática de la Universidad Autónoma de Madrid (ed.): *Corpus Oral de Referencia de la Lengua Española Contemporánea (CORLEC). Ejemplos*. Online Publikation: <http://www.llf.uam.es/ESP/Info%20Corlec.html> (10.03.2017).

⁴⁴⁸ Vgl. Laboratorio de Lingüística Italiana (ed.): *Corpus Didáctico Anotado Lingüísticamente (C-Or-DiAL)*. Online Publikation: <http://lablita.dit.unifi.it/corpora/cordial/> (19.03.2017).

aufgenommen – ohne vorherige Kenntnis der Sprecher, jedoch mit späterer Einverständniserklärung.⁴⁴⁹

Das Korpus ist im Internet kostenlos und frei zugänglich (<http://lablita.it/app/cordial/corpus.php>), alle Aufnahmen sind als Audiodateien im *wav*-Format herunterzuladen und die dazugehörigen Transkriptionen als *txt*-Dateien.⁴⁵⁰ Mittels einer Suchmaske können die Dateien nach Gesprächssituation, didaktischem Niveau des *Gemeinsamen Europäischen Referenzrahmens für Sprachen* (siehe Kapitel 3.1.2.), Schlagwörtern und kommunikativer Funktion durchforstet werden, die Suche nach einzelnen Lexemen ist jedoch nicht möglich.⁴⁵¹ Außerdem gibt es eine Übersicht mit allen Aufnahmen des Korpus, die zusätzliche Informationen zum Thema und zur Länge des Texts (sowohl zur Anzahl der Lexeme als auch zur Dauer in Minuten) sowie zur Anzahl der Sprecher und dem didaktischen Niveau bietet.⁴⁵² Zudem ist das Korpus als CD-ROM verfügbar.⁴⁵³

In den orthographischen Transkriptionen sind prosodische Einheiten und andere Züge der mündlichen Ausdrucksweise, wie Reformulierungen, annotiert, wobei hierfür die *tags* des C-ORAL-ROM verwendet wurden (siehe Kapitel 2.4.2.2.). Auch kommunikative Funktionen wurden anhand einer vereinfachten Version der Richtlinien des *Instituto Cervantes* annotiert. Alle Aufnahmen zeichnen sich durch ein mehr oder weniger informelles Register und einen hohen Grad an Spontaneität aus. Dennoch werden sie in zwei Gruppen eingeteilt: Zum einen in spontane Dialoge und Konversationen, zum anderen in nicht spontane und öffentliche Interviews, Unterrichtsstunden, Konferenzen und diverse Gespräche.⁴⁵⁴

⁴⁴⁹ Vgl. Laboratorio de Lingüística Italiana (ed.): *Corpus Didáctico Anotado Lingüísticamente (C-Or-DiAL)*. Op. cit. (19.03.2017).

⁴⁵⁰ Vgl. Laboratorio de Lingüística Italiana (ed.): *Corpus Didáctico Anotado Lingüísticamente (C-Or-DiAL)*. Op. cit. (19.03.2017).

⁴⁵¹ Vgl. Laboratorio de Lingüística Italiana (ed.): *Acceso a las sesiones de C-Or-DiAL. Búsqueda avanzada*. Online Publikation: <http://lablita.dit.unifi.it/app/cordial/search.php> (19.03.2017).

⁴⁵² Vgl. Laboratorio de Lingüística Italiana (ed.): *Acceso a las sesiones de C-Or-DiAL. Indices de todo el Corpus*. Online Publikation: <http://lablita.dit.unifi.it/app/cordial/corpus.php> (19.03.2017).

⁴⁵³ Vgl. Nicolás Martínez, Carlota (2012): *Corpus C-Or-DiAL (Corpus Oral Didáctico Anotado Lingüísticamente)*. Madrid: Liceus.

⁴⁵⁴ Vgl. Laboratorio de Lingüística Italiana (ed.): *Corpus Didáctico Anotado Lingüísticamente (C-Or-DiAL)*. Op. cit. (19.03.2017).

2. Kategorisierung der Korpora

@Archivos: dial_32_A_DENTELLADAS.txt, dial_32_A_DENTELLADAS.wav
@Título: a dentelladas
@Participantes: CAR, Carlota (mujer, C, 3, profesora, Madrid, vive en Italia desde hace más de 20 años)
MIG, Miguel (hombre, C, 3, investigador, Madrid)
@Relación entre los participantes: MIG y CAR son amigos
@Situación: comiendo en casa de MIG
@Tema: MIG hace suposiciones sobre las dificultades en las relaciones futuras entre los que ahora son jóvenes adolescentes y los hablantes, MIG no tiene hijos y CAR es madre de Valeria y Alberto
@Palabras clave: juventud, familia, gustos, futuro
@Uso didáctico: B1
@Nivel de comprensión del texto: B1
@Palabras nuevas: de lejos, patadas, coñazo, a dentelladas, cuchara, órdago, enfoque, tener pinta, dar para el pelo, enfoque, cotizar, seguridad social
@Funciones comunicativas: 1.2 pedir información, 1.3 dar información, 2.2 dar una opinión, valorar, 2.13 expresar certeza y evidencia, falta de certeza y evidencia, posibilidad, 3.1 preguntar por gustos e interés y expresar gustos e intereses, 3.7 preguntar por deseos y expresar deseos, 4.26 preguntar por obligación y necesidad y expresar obligación y necesidad o falta de obligación o de necesidad, 6.16 introducir palabras de otros y citar
@Observaciones lingüísticas: baja velocidad de elocución; oraciones temporales; riqueza verbal; tiempo futuro; citas
@Duración y número de palabras: 00:05:16 – 670
@Transcriptores y revisores: Carlota Nicolás Martínez, Riccardo Saracinelli, Edoardo Battistini
@Grabación original: 32_MIGUEL_SOBREMESA, 00:55:01, Madrid, 2007
*CAR: bueno / ya veremos / como sigue //
*MIG: continuará //
*CAR: continuará //
*MIG: sí sí a mí me encanta / esto de la sociología barata / es un tema que me encanta / pero de lejos //
*CAR: yyy.
*MIG:/ mira mira los niños mira que graciosos / vale / hola //
*CAR: yyy.
*MIG:/ pero de cerca / vamos ... o sea que sí / que estoy deseando ver cómo tienen los niños / gente de la generación de Valeria / o de / a ver a ver qué hacéis vosotros / que cuenten que digan //
*CAR: como curiosidad sociológica //

Abbildung 43: C-Or-DiAL – Transkription eines Dialogs zu den Themen Familie, Zukunft, Vorlieben und Jugend.⁴⁵⁵

Der obige Ausschnitt lässt bereits erkennen, dass die Metadaten der einzelnen Aufnahmen sehr umfassend sind. Zu Beginn jeder Datei werden nicht nur die Sprecher, deren Beziehung zueinander, das Thema, zugehörige Schlüsselwörter sowie das didaktische Niveau angegeben, sondern auch neues Vokabular, die darin enthaltenen kommunikativen Funktionen oder linguistische Beobachtungen. An die Metadaten schließt sich die Transkription der eigentlichen Aufnahme an.

2.4.3. Multimodale Korpora

Neben den Korpora gesprochener und geschriebener Sprache gibt es noch die multimodalen Korpora. Hier bestehen die Primärdaten aus einer zusätzlichen Videospur, was beispielsweise für die Gestenforschung oder für die Gebärdensprache von Bedeutung ist.⁴⁵⁶

⁴⁵⁵ Laboratorio de Lingüística Italiana (ed.): *Acceso a las sesiones de C-Or-DiAL. Indices de todo el Corpus*. Op. cit. (19.03.2017).

⁴⁵⁶ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 105.

2.4.3.1. *Sacodeyl Search Tool*

Ein spanisches Beispiel für ein multimodales Korpus ist das *Sacodeyl Search Tool* (kurz: *Sacodeyl*), ein pädagogisches Korpus in sieben europäischen Sprachen (Deutsch, Englisch, Französisch, Italienisch, Litauisch, Rumänisch und Spanisch). Es ist im Internet frei zugänglich (<http://www.um.es/sacodeyl/>) und hält Videosequenzen von Schülern zwischen 13 und 18 Jahren zu Alltagsthemen bereit. Die Interviews wurden pädagogisch annotiert und für den Sprachlernprozess aufbereitet.⁴⁵⁷

Das Ziel dieses Projekts ist die pädagogische, sprachlernorientierte Nutzung eines mündlichen *Teen-Talk*-Korpus. Hierfür wurden bereits existierende *ICT-Tools* (*Information and Communications Technology*) für die Integration von Korpora und korpusbasierten Methoden des *Data Driven Learning* (DDL) in den gewöhnlichen Sprachunterricht pädagogisch angepasst. Mithilfe dieses Korpus sollen junge Fremdsprachenlerner in Europa die Sprache selbst erforschen und sich selbstständig mit Lernaufgaben einer besonderen Art beschäftigen.⁴⁵⁸

Jedes Subkorpus enthält 20 bis 25 Videos von jeweils 10 Minuten Länge. Die Interviews wurden entweder mit einzelnen Schülern oder mit Schülerpaaren durchgeführt. Um die Subkorpora thematisch vergleichbar zu gestalten, wurde ein allgemeiner Fragenkatalog entwickelt, der eine Fülle von Themen abdeckt, wie beispielsweise die eigene Person, Heim und Familie, Hobbys und Interessen sowie Urlaub. Zudem wurden die Fragen so gestellt, dass sie den Interviewten Antworten in bestimmten *Proficiency Levels* des *Common European Frame of Reference* (siehe auch Kapitel 3.1.2.) als auch pädagogisch relevante Sprachstrukturen entlocken sollten.⁴⁵⁹

Im *Sacodeyl* kann der Nutzer entweder durch die Interviews browsen und sich das ein oder andere interessante Interview ansehen beziehungsweise durchlesen, er kann das Korpus aber auch gezielt durchsuchen. Eine Möglichkeit ist die Suche nach Kategorien wie Thema, Grammatik, Lexik, Varietät/Stil oder Sprachniveau. Zudem kann nach ein

⁴⁵⁷ Vgl. Pérez-Paredes, Pascual: *Sacodeyl. European Youth Language*. Online Publikation: <http://sacodeyl.inf.um.es/sacodeyl-search2/> (18.11.2016).

⁴⁵⁸ Vgl. Pérez-Paredes, Pascual: *Sacodeyl. Project Overview*. Online Publikation: <http://www.um.es/sacodeyl/> (24.01.2017).

⁴⁵⁹ Vgl. Kohn, Kurt/ Hoffstaedter, Petra (2008): *SACODEYL Corpus-Based Language Learning – Guidelines for Teachers and Learners*. Online Publikation: http://www.um.es/sacodeyl/data/publications/SACODEYL_guidelines_for_teachers.pdf (10.01.2017), S. 5-6.

2. Kategorisierung der Korpora

oder mehreren Lexemen gleichzeitig gesucht werden und auch eine Frequenzliste ist verfügbar.

Left Context	Match	RightContext
un poco y eso es lo que me	<i>gustaría</i>	
Francia y Lon y Inglaterra si me me	<i>gustaría</i>	, me llamaría la atención . ¿Por el idioma y
sé, yo cuando cuando termine Bachillerato me me	<i>gustaría</i>	, porque me gustan mucho las ciencias , hacer Ingeniería
he salido fuera de España pero sí me	<i>gustaría</i>	, sí
he tenido varios gatos. Sobre todo pues me	<i>gustaría</i>	, también me gusta mucho la jardinería tengo un
el periodismo, algo de letras, y sí me	<i>gustaría</i>	, y como mis tías dos de mis tías
verlo, en Africa algún lugar de éstos me	<i>gustaría</i>	, y de mayor ir a ayudar a alguna
sea España? O sea, a Italia. Sí, me	<i>gustaría</i>	. No sé , no sé que es un país
he llegado a arrepentir pero ahora sí me	<i>gustaría</i>	a lo mejor saber tocar algún instrumento , alguna
la jardinería tengo un jardín pequeño y me	<i>gustaría</i>	agrandarlo y también me gusta cocinar pero tengo
y además me gusta aprender idiomas y me	<i>gustaría</i>	aprender el italiano y además visitar Italia porque
aparte de que tengo que decir que me	<i>gustaría</i>	en todos , por supuesto no rechazo nunca un
vamos leer no, pero lo que más me	<i>gustaría</i>	es salir y to eso , viajar también me
sí vivo es durante poco tiempo pero me	<i>gustaría</i>	eso para conocer a gente para hacerte amigo
hacer, bueno yo querría ser médico y me	<i>gustaría</i>	estudiar un año medicina en Italia porque me
estudiando, trabajar o ? Sí sí, a mí me	<i>gustaría</i>	estudiar y sacarme una carrera y , pues eso

Abbildung 44: *Sacodeyl* – Anzeige der Konkordanzen für *gustaría*.⁴⁶⁰

Bei der Suche nach einem Begriff werden die Treffer wie gewöhnlich als Konkordanzen angezeigt und der Kontext so geordnet, dass der Suchbegriff zentral steht und hervorsteht.

2.4.3.2. *Backbone Spanish*

Backbone ist ein europäisches Projekt, das vom *Lifelong Learning Project* der *European Commission* finanziert wurde und acht Partner aus sieben europäischen Ländern miteinbezieht. Das *Backbone*-Projekt entwickelte webbasierte, pädagogische Korpora von aufgezeichneten Interviews mit englischen, französischen, deutschen, polnischen, spanischen und türkischen Muttersprachlern sowie Sprechern des Englischen als *lingua franca*. Die Interviews beziehen sich auf Themen wie Kultur, Bildung, Gesundheit, Wirtschaft und Technologie. Die Interviewtranskripte sind pädagogisch annotiert im Hinblick auf thematische, grammatikalische sowie lexikalische Eigenschaften, die für das Lehren und Lernen von Sprachen relevant sind.⁴⁶¹ Die einzelnen Interviews sind zudem in kleine Sequenzen eingeteilt, die mit einem Titel versehen sind. Auf diese Weise kann sich der Nutzer einen Überblick über den Inhalt des Videos verschaffen, ohne sich das ganze Transkript durchlesen zu müssen.

⁴⁶⁰ Pérez-Paredes, Pascual: *Sacodeyl. European Youth Language. Search ,gustaría'*. Online Publikation: <http://sacodeyl.inf.um.es/sacodeyl-search2/faces/search.jsp> (11.03.2017).

⁴⁶¹ Vgl. Kohn, Kurt: *Backbone Project*. Online Publikation: <http://projects.ael.uni-tuebingen.de/backbone/moodle/> (15.02.2018).

Das Korpus ist im Internet frei zugänglich (<http://webapps.ael.uni-tuebingen.de/backbone-search/faces/search.jsp>) und enthält 25 Videointerviews von Sprechern unterschiedlicher sozialer Schichten, die nach thematischen und linguistischen Kategorien durchsucht werden können. Es ist ähnlich aufgebaut wie das Korpus *Sacodeyl* (welches außerdem in der Korpusdatenbank des *Backbone*-Projekts aufgelistet und verlinkt ist), und die Videos können ebenso im Bereich *Section Search* nach Themen, Grammatik, Lexik, kommunikativen Funktionen oder Sprachniveau gefiltert werden. Die Anzeige von Konkordanzen mit bis zu drei Suchbegriffen ist im Bereich *Concordances* möglich, und im Bereich *Co-Occurrence* kann nach einzelnen Lexemen gesucht werden, wobei hier keine Konkordanzen als Ergebnis angezeigt werden, sondern ganze Textabschnitte, welche zudem als Textdatei exportiert werden können.⁴⁶² Anhand dieser Option kann ermittelt werden, in welchen Videos ein bestimmter Begriff vorkommt, um zu sehen, welche Interviews für ein bestimmtes Themengebiet geeignet sind. Die folgende Abbildung enthält einen Abschnitt aus einer exportierten Textdatei, die im Anschluss an die Suche nach dem Substantiv *deporte* im Bereich *Co-Occurrence* erstellt wurde:

María: Muy bien. ¿Y conoces si se hace algo parecido en Murcia? Esto lo haces en Alicante, ¿sabes si en Murcia se fomenta?

Sabina: Sí, en Murcia hay '*Deporte* en la calle' hay que eso es un día de *deporte* en la calle en el que se prepara como una ins mini instalaciones de mini-tenis, se dibuja una pista de tenis en el suelo, unas pistas de mini-basket de fútbol y se practican los deportes en la calle. Eso es una forma de fomentar y atraer a la gente a los centros deportivos, a los pabellones municipales. Es una forma de acercar el *deporte* a la gente que a lo mejor no conoce o que no le ha llegado información por algún tipo a casa o algo de eso. De esa forma se puede ir metiendo a la gente un poquitín en el mundo del *deporte*.

Abbildung 45: *Backbone Spanish* – Suche nach dem Substantiv *deporte* im Bereich *Co-Occurrence*.⁴⁶³

Die Videos mit einer Länge zwischen zehn und zwanzig Minuten können auch nach diatopischen Varietäten ausgewählt werden. So stehen Interviews mit dem Akzent von *Aragón*, *Andalucía*, *Castilla-La Mancha*, *Murcia*, *Comunidad Valenciana*, *Galicia* oder *Cantabria* sowie in neutralem Spanisch zur Verfügung.

Auf der Internetseite des Korpus *Backbone Spanish* werden zudem didaktisch aufbereitete Materialien beispielsweise zur Erarbeitung des Themas *Mundo del Trabajo*

⁴⁶² Vgl. Kohn, Kurt: *Backbone*. Online Publikation: <http://webapps.ael.uni-tuebingen.de/backbone-search/faces/search.jsp> (15.02.2018).

⁴⁶³ Kohn: *Backbone*. Op. cit. (15.02.2018).

zur Verfügung gestellt. Bei den Aufgaben zum Hörverstehen fehlen allerdings die Audiodateien, sodass diese Aufgaben, welche dazu vorgesehen sind, von den Lernern selbstständig am Computer bearbeitet zu werden, nicht genutzt werden können.

2.4.3.3. SPT – *Spanish Corpus Proficiency Level Training*

Das *Spanish Corpus Proficiency Level Training* (kurz: SPT) ist ein multimodales Lernerkorpus aus Texten von Fremdsprachenlernern mit unterschiedlichem Leistungsstand, welches mit der Zielsetzung erstellt wurde, Lehrkräfte für die Bewertung von Schülerantworten zu trainieren. Neben Lehrern können auch fortgeschrittene Spanischler das SPT zur Verbesserung ihrer Sprachkenntnisse nutzen. Die Homepage enthält Beschreibungen der einzelnen Sprachniveaus, Videos und Übungen zur Einstufung der Lernertexte. Dabei gibt es Übungsfragen zu den Bereichen Vokabular, Grammatik, Flüssigkeit und Gesprächsführung.⁴⁶⁴

The screenshot displays the SPT interface with a navigation bar at the top containing tabs for 'grammar accuracy', 'vocabulary', 'fluency', 'conversation management', 'generalizations', and 'evaluate'. The 'grammar accuracy' tab is selected. Below the tabs, there are four numbered questions with multiple-choice options (a-e). Questions 1 and 3 focus on grammatical control and overall understanding, while questions 2 and 4 focus on grammatical inaccuracies and correction strategies. At the bottom of the question area, there are two buttons: 'Score' and 'Try Again'.

grammar accuracy vocabulary fluency conversation management generalizations evaluate

1. In what areas did this learner show fairly accurate grammatical control?

- a. ☐ Subject-verb agreement
- b. ☐ Gender-number agreement
- c. ☐ Subjunctive Use
- d. ☐ Verb conjugations
- e. ☐ None of the above

2. Generally, what were the main areas in which this learner showed some grammatical inaccuracies?

- a. ☐ Subject-verb agreement
- b. ☐ Noun-adjective agreement
- c. ☐ Subjunctive Use
- d. ☐ Verb conjugations
- e. ☐ None of the above

3. Which errors were most glaring in terms of overall understanding of the learner's response?

- a. ☐ Subject-verb agreement
- b. ☐ Noun-adjective agreement
- c. ☐ Subjunctive Use
- d. ☐ Verb conjugations
- e. ☐ None of the above

4. How did the learner often correct his/her grammar mistakes, if at all?

- a. ☐ Self-correction without asking for help
- b. ☐ Self-correction after requesting help
- c. ☐ By repeating what the teacher asked him/her to say.
- d. ☐ No correction

Score Try Again

Abbildung 46: SPT – Fragen zur Grammatik in einem Lernervideo.⁴⁶⁵

Derzeit verfügt das Korpus über Aufnahmen von 16 Spanischlernern mit englischer Muttersprache, welche sich zu den Themen *Hometown, Past, Sports, Politics, Other Countries, Family, Comparisons, Major* (Studienhauptfach) und *Apartment* in verschiedenen Videos von jeweils ein bis fünf Minuten Länge äußern. Der Nutzer kann die Videos zu verschiedenen Lernern und Themen auswählen sowie den jeweiligen Sprecher, nachdem er ein paar Videos von ihm gesehen hat, bewerten.⁴⁶⁶

⁴⁶⁴ Vgl. Koike, Dale: *Spanish Corpus Proficiency Level Training. Intro*. Online Publikation: <http://www.laits.utexas.edu/spt/intro> (05.10.2017).

⁴⁶⁵ Koike, Dale: *Spanish Corpus Proficiency Level Training. Video Activity*. Online Publikation: <http://www.laits.utexas.edu/spt/video> (05.10.2017).

⁴⁶⁶ Vgl. Koike: *Spanish Corpus Proficiency Level Training. Video Activity*. Op. cit. (05.10.2017).

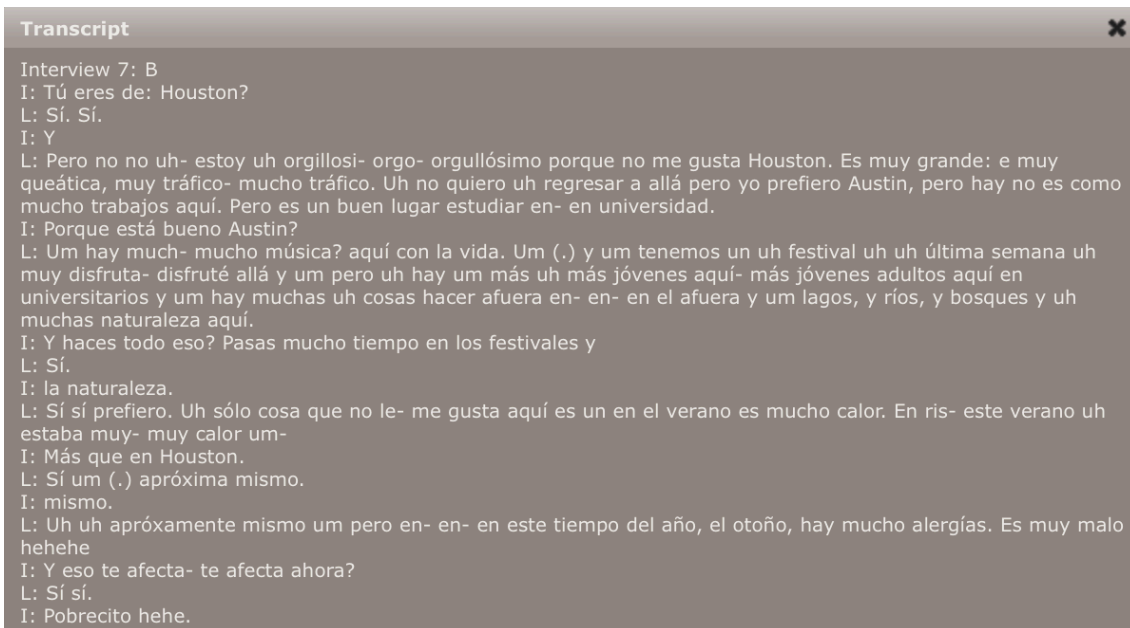


Abbildung 47: SPT – Transkript eines Lernervideos.⁴⁶⁷

Zudem kann man sich, wie in obiger Abbildung zu sehen ist, ein Transkript des jeweils ausgewählten Videos anzeigen lassen. Das Korpus ist im Internet kostenlos verfügbar und ohne Anmeldung zugänglich (<http://www.laits.utexas.edu/spt/intro>).

2.5. Annotation

Bei diesem Kriterium muss zunächst festgestellt werden, ob ein Korpus überhaupt annotiert ist. Korpora beziehungsweise deren Texte können in Bezug auf die Bereiche Morphosyntax, Syntax, Semantik, Pragmatik, Fehler, aber auch Textstruktur, Informationsstruktur, Phonetik/Prosodie oder Gestik annotiert sein, wobei die morphosyntaktische und die syntaktische Annotation am häufigsten vertreten sind.⁴⁶⁸

2.5.1. Morphosyntaktische Annotation

Die morphosyntaktische Annotation, auch *POS-tagging* genannt, ist die gebräuchlichste Form der Annotation. Zudem gibt es die Annotation von Flexionsmorphologie, welche häufig auch unter der Bezeichnung ‚morphologische Annotation‘ oder ‚Lemmatisierung‘ zu finden ist. Hier wird, wie bereits in Kapitel 1.4.5. beschrieben, das *token* (die einzelne Wortform, beispielsweise *libros*) analysiert und auf

⁴⁶⁷ Koike: *Spanish Corpus Proficiency Level Training. Video Activity*. Op. cit. (05.10.2017).

⁴⁶⁸ Vgl. Becker (2013): op. cit., S. 193-194.

2. Kategorisierung der Korpora

seine Grundform, das Lemma (hier: *libro*), zurückgeführt. Die Flexionsmorphologie umfasst dabei die Kategorien Kasus, Genus, Numerus, Person, Tempus und Modus.⁴⁶⁹

Unter den spanischen Korpora zeichnen sich die beiden Teile des CdE (siehe Kapitel 2.1.1.3., 2.2.1.3. und 2.6.1.) sowie das CORPES (siehe Kapitel 2.1.1.2.) besonders aus. Die Annotation des CdE umfasst dabei die Angabe der Lemmata sowie die morphosyntaktischen Kategorien.⁴⁷⁰ Es ist also nach Wortarten getaggt und verfügt über flexionsmorphologische Informationen. Auch das CORPES ist morphosyntaktisch annotiert und lemmatisiert.⁴⁷¹

Im CORPES kann sich der Nutzer bei allen Treffern in der Anzeige der Konkordanzen die zugehörigen Informationen anzeigen lassen, indem er die Maus über das jeweilige Lexem bewegt. Bei der Suche nach dem Lemma *libro* in der Form *libros* wird bei einer Konkordanz beispielsweise folgende Information angegeben:

libro - sustantivo masculino plural común
Cuando termino un libro procuro no tener expectativas respecto a él, estoy convencida de que los libros tienen su destino y yo no puedo más que ser una espectadora más de lo que va a pasar con él.

Abbildung 48: CORPES – Flexionsmorphologische Informationen zum Lemma *libro* in der Form *libros*.⁴⁷²

Neben den erwähnten, umfassenden Korpora CdE und CORPES gibt es folgende kleinere, morphosyntaktisch annotierte Korpora: das *Corpus Tècnic del IULA* und das Korpus *El Grial*.

2.5.1.1. CT-IULA – *Corpus Tècnic del IULA*

Ein morphosyntaktisch annotiertes und lemmatisiertes Korpus der spanischen Sprache ist das *Corpus Tècnic* des *Institut de Lingüística Aplicada de la Universitat Pompeu Fabra (IULA de la UPF)*, welches auf deren Internetseite (<http://bwananet.iula.upf.edu>) kostenlos zur Verfügung gestellt wird und mittels der Plattform *BwanaNet* die Durchführung verschiedener Suchoptionen erlaubt.⁴⁷³ Zunächst muss die Sprache der Dokumente ausgesucht werden, wobei der Nutzer die Wahl

⁴⁶⁹ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 70.

⁴⁷⁰ Vgl. Becker (2013): op. cit., S. 193-194.

⁴⁷¹ Vgl. Real Academia Española (ed.): *CORPES*. Op. cit. (24.10.2016).

⁴⁷² Real Academia Española (ed.): *CORPES. Buscar 'libros'*. Online Publikation: <http://web.frl.es/CORPES/org/publico/pages/consulta/entradaCompleja.view> (11.03.2017).

⁴⁷³ Vgl. Institut Universitari de Lingüística Aplicada (ed.): *BwanaNet. Programa de explotación del Corpus Técnico de IULA*. Online Publikation: <https://bwananet.iula.upf.edu> (26.10.2016).

zwischen Katalanisch, Spanisch und Englisch hat. Es verfügt über 1813 spanische Texte mit mehr als 33 Mio. Lexemen, 1191 katalanische (26 Mio. Lexeme) und 535 englische Texte (3,5 Mio. Lexeme). Im nächsten Schritt können die zu durchsuchenden Dokumente nach dem Thema oder der Größe ausgewählt oder eine zuvor gespeicherte Sitzung wiederaufgenommen werden (Option a). Es stehen Themengebiete wie Informatik, Umwelt, Recht, Medizin, Wirtschaft Linguistik und Allgemeines zur Auswahl, es kann allerdings auch im ganzen Korpus gesucht werden (Option b).⁴⁷⁴

Das Korpus bietet verschiedene Suchoptionen, so kann man sich beispielsweise die *types*, *tokens* und Wortarten des Korpus der Häufigkeit nach anzeigen lassen, eine Suche nach einfachen Konkordanzen auswählen oder eine Standard- beziehungsweise komplexe Suche anhand der Eingabemaske initialisieren. Das Korpus lässt bei der Standardsuche nach Konkordanzen beispielsweise die Angabe von fünf oder mehr Einheiten zu, wobei auch die Wortart angegeben werden kann. Dadurch, dass der Nutzer unter anderem die Möglichkeit hat, ausschließlich in Texten zu bestimmten Themengebieten zu suchen, kann die Trefferzahl von vornherein eingeschränkt werden.

<s>Para disminuir la	contaminación	a l mínimo , analizar primero las muestras diluidos .</s>
es , a l mismo tiempo , la mejor solución para combatir la	contaminación	
aíses en los se han realizado programas de lucha contra la	contaminación	de las aguas , estos programas han conducido siempre a l
s que van de la mano : disminuir la demanda y eliminar la	contaminación	.</s>
técnicas destinadas a l ahorro de agua reducen no sólo la	contaminación	sino también los mayores peligros que conlleva la irrigació
<s>Esto supone menos de l 10 por 100 de los consumos y	contaminaciones	urbanas cuyas cifras van según los datos de la demanda p
<s>La	contaminación	difusa por el alcantarillado no se conoce , aunque puede s
puede ser grande , sin embargo sí que se ha estudiado la	contaminación	de la escorrentía pluvial en las ciudades .</s>
esto no permite el funcionamiento de las depuradoras y la	contaminación	después de las lluvias puede ser muy grande , a la vez qu
<s>La gran	contaminación	de las aguas de la escorrentía pluvial y el problema de las

Abbildung 49: CT-IULA – Suche nach dem Lemma *contaminación* im Unterbereich *Ciencia y tecnología ambiental y energética*.⁴⁷⁵

Die Suche nach dem Lemma *contaminación* im Unterbereich *Ciencia y tecnología ambiental y energética* des Themengebiets *Medio Ambiente* liefert beispielsweise 126 Konkordanzen in 19 Dokumenten. Auf diese Weise können die gesuchten Begriffe gezielt im jeweiligen Kontext gesucht werden, um konkrete und nicht nur allgemeine Ergebnisse zu erzielen. Die Lemmatisierung ermöglicht in diesem Fall zudem die Anzeige der Treffer im Singular und Plural.

⁴⁷⁴ Vgl. Institut Universitari de Lingüística Aplicada (ed.): *BwanaNet. Programa de explotació del Corpus Técnico de IULA*. Op. cit. (26.10.2016).

⁴⁷⁵ Institut Universitari de Lingüística Aplicada (ed.): *BwanaNet. Programa de explotació del Corpus Técnico de IULA. Buscar ,contaminación' en ,Ciencia y tecnología ambiental y energética'.* Online Publikation: <http://bwananet.iula.upf.edu/cgi-bin/bwananet0/viewSimple.pl> (13.03.2017).

2. Kategorisierung der Korpora

2.5.1.2. El Grial

Ein weiteres morphosyntaktisch annotiertes und lemmatisiertes Korpus ist das Korpus *El Grial* der *Pontificia Universidad Católica de Valparaíso*. Sie bietet auf ihrer Internetseite verschiedene Suchmöglichkeiten im Korpus *El Grial*, welches 35 Dokumente mit insgesamt 5,4 Mio. Lexemen umfasst. Zunächst kann zwischen einer einfachen Suche, einer komplexen Suche, dem *Manchador de Textos* und der *Visualización Output* gewählt werden. Bei allen Suchanfragen muss der Nutzer zunächst ein Korpus auswählen. Hierbei können die Texte nach Modus, Thema, Textsorte, Register, Korpus und Dokument zusammengestellt werden, wobei im Hinblick auf den Modus zu erwähnen ist, dass ausschließlich schriftliche Dokumente zur Verfügung stehen.⁴⁷⁶

Bei der einfachen Suche (*Búsqueda Simple*) kann der Nutzer Informationen zu Wortform, Lemma, Wortart, Genus und Numerus angeben und sich die Ergebnisse als Konkordanzen anzeigen lassen. Hierbei liegt die Grundeinstellung bei bis zu 5000 Konkordanzen mit jeweils fünf Lexemen vor und nach dem Suchbegriff. Es können jedoch auch alle Konkordanzen auf einmal angezeigt werden, und die Spanne kann auf bis zu 15 Lexeme ausgeweitet werden. Zudem kann ausgewählt werden, dass der komplette Kontext, also der vollständige Satz, in dem sich der Suchbegriff befindet, angezeigt werden soll.⁴⁷⁷

Cantidad de resultados: 86

LÍNEA	IZQUIERDA	BÚSQUEDA	DERECHA
1	y restaurantes y confiterías , y luego a los ómnibus (donde ahora hay también	televisión) y a los taxis y finalmente a las calles . Debo decir que jamás
2	. Aunque no podría evitar se la curiosidad de los científicos ni las cámaras de	televisión	, a l menos se podría detener la invasión de aventureros y colonos dispuestos a
3	una breve nota , a una columna , en página interior . Los noticieros de	televisión	, salvo el primer día , olvidaron el asunto y se dedicaron a cosas más
4	no soportarían mi presencia más de ochohoras seguidas , son los escritores - presentadores de	televisión	, los que adoran la locura de Barrendoáin mientras mueven sensatamente la cabeza , y
5	brasileñas . Finalmente , durante las largas horas de tedio en su pequeño cuarto sin	televisión	, María concluyó : a) Nunca llegaría a encontrar lo que estaba buscando ,
6	y compartían con énfasis adolescente los ídolos de l cine , la canción , la	televisión	, los locales de baile y los bares , los cafés universitarios , las primeras
7	días siguientes , las contradicciones de l proceso , como decían los sociólogos en la	televisión	, se hicieron sentir en la caleta de manera más rigurosa que retórica . Los

Abbildung 50: *El Grial* – Ergebnisanzeige bei der Suche nach *televisión* in der Option *Búsqueda Simple*.⁴⁷⁸

Die Einstellungsoptionen zu den Konkordanzen sind bei der komplexen Suche (*Búsqueda Compleja*) die gleichen, doch hier kann der Nutzer nicht nur einen Suchbegriff

⁴⁷⁶ Vgl. Pontificia Universidad Católica de Valparaíso (ed.): *El Grial. Interfaz de etiquetaje e interrogación de corpus textuales*. Online Publikation: <http://www.elgrial.cl> (26.10.2016).

⁴⁷⁷ Vgl. Pontificia Universidad Católica de Valparaíso (ed.): *El Grial*. Online Publikation: <http://www.elgrial.cl/index2015.php> (26.10.2016).

⁴⁷⁸ Pontificia Universidad Católica de Valparaíso (ed.): *El Grial*. Op. cit. (26.10.2016).

genauer spezifizieren, sondern nach einer Sequenz von bis zu 6 *tokens* in einer Reihe suchen, bei denen stets Informationen zu Wortform, Lemma, Wortart, Genus und Numerus angegeben werden können:

Extensión del corpus seleccionado: 5392544 tokens.

Mostrar						
<input checked="" type="radio"/> Mostrar hasta <input type="text" value="5000"/> contextos en pantalla <input type="radio"/> mostrar todos los contextos			<input type="radio"/> Contexto completo (oracional) <input checked="" type="radio"/> Contexto parcial <input type="text" value="5"/> Unidades lqz y Der.			
<input checked="" type="checkbox"/> Contextos ordenados						

Modo						
Secuencia	Token 1	Token 2	Token 3	Token 4	Token 5	Token 6
Despliegue	<input checked="" type="radio"/> Forma <input type="radio"/> POS <input type="radio"/> FDP	<input checked="" type="radio"/> Forma <input type="radio"/> POS <input type="radio"/> FDP	<input checked="" type="radio"/> Forma <input type="radio"/> POS <input type="radio"/> FDP	<input checked="" type="radio"/> Forma <input type="radio"/> POS <input type="radio"/> FDP	<input checked="" type="radio"/> Forma <input type="radio"/> POS <input type="radio"/> FDP	<input checked="" type="radio"/> Forma <input type="radio"/> POS <input type="radio"/> FDP
Forma	<input type="text" value="="/>	<input type="text" value="="/>	<input type="text" value="="/>	<input type="text" value="="/>	<input type="text" value="="/>	<input type="text" value="="/>
Lema	<input type="text" value="="/>	<input type="text" value="="/>	<input type="text" value="="/>	<input type="text" value="="/>	<input type="text" value="="/>	<input type="text" value="="/>
POS	seleccionar	seleccionar	seleccionar	seleccionar	seleccionar	seleccionar
Genus	seleccionar	seleccionar	seleccionar	seleccionar	seleccionar	seleccionar
Número	seleccionar	seleccionar	seleccionar	seleccionar	seleccionar	seleccionar
Dep. Func.	seleccionar	seleccionar	seleccionar	seleccionar	seleccionar	seleccionar

Abbildung 51: *El Grial* – Benutzeroberfläche der Option *Búsqueda Compleja*.⁴⁷⁹

Mittels der Option *El Manchador de Textos* kann der Nutzer gleichzeitig nach verschiedenen Sequenzen mehrerer *tokens* suchen und sich diese farblich getrennt in längeren Textausschnitten anzeigen lassen. Zu Beginn der Ergebnissseite wird die gesamte Trefferanzahl jeder einzelnen Sequenz zusammen mit der zugewiesenen Farbe angezeigt. Im folgenden Beispiel wurde nach *para* mit einem Infinitiv und *para que* gefolgt von einer Verbform im *subjuntivo* gesucht:

Documento 1: CLL-2016-02	
<p>La casa de los espíritus Isabel Allende Rosa, la bella Capítulo I Barrabás llegó a la familia por vía marítima, anotó la niña Clara con su delicada caligrafía. Ya entonces tenía el hábito de escribir las cosas importantes y más tarde, cuando se quedó muda, escribía también las trivialidades, sin sospechar que cincuenta años después, sus cuadernos me servirían</p> <p>para rescatar la memoria de lo pasado y para sobrevivir a mi propio espanto. El día que llegó Barrabás era jueves Santo. Venía en una jaula indigna, cubierto de sus propios excrementos y orines, con una mirada extraviada de preso miserable e indefenso, pero ya se adivinaba - por el porte real de su cabeza y el tamaño de su esqueleto - el gigante legendario que llegó a ser. Aquél era un día aburrido y otoñal, que en nada presagiaba los acontecimientos que la niña escribió para que fueran recordados y que ocurrieron durante la misa de doce, en la parroquia de San Sebastián, a la cual asistió con toda su familia. En señal de duelo, los santos estaban tapados con trapos morados, que las beatas desempolvaban anualmente de lo ropero de la sacristía, y bajo las sábanas de luto, la corte celestial parecía un amasijo de muebles esperando la mudanza, sin que las ve las, el incienso o los gemidos de lo órgano, pudieran contrarrestar ese lamentable efecto. Se erguían amenazantes bultos oscuros en el lugar de los santos de cuerpo entero, con sus rostros idénticos de expresión constipada, sus elaboradas pelucas de cabello de muerto, sus rubies, sus perlas, sus esmeraldas de vidrio pintado y sus vestuarios de nobles florentinos. El único favorecido con el luto era el patrono de la iglesia, san Sebastián, porque en Semana Santa le ahorra a los fieles el espectáculo de su cuerpo torcido en una postura indecente, atravesado por media docena de flechas, chorreando sangre y lágrimas, como un homosexual sufriente, cuvas llagas, milagrosamente frescas gracias a lo pincel de lo padre Restrepo.</p>	<p>Resultados:</p> <p>Secuencia 1: 749 casos</p> <p>Secuencia 2: 48 casos</p> <p>Casos encontrados: 797 Tipos encontrados: 2 Tokens texto: 190645 IDL: (797*2/190645) 0.0083611</p> <p>Enlaces: Secuencia: 1 Secuencia: 2 </p> <p>Secuencia 1: 1 2 3 4 5 6 7 8 9 10 11</p>

Abbildung 52: *El Grial* – Option *El Manchador de Textos* bei der Suche nach *para* + *infinitivo* und *para que* + *subjuntivo*.⁴⁸⁰

⁴⁷⁹ Pontificia Universidad Católica de Valparaíso (ed.): *El Grial*. Op. cit. (26.10.2016).

⁴⁸⁰ Pontificia Universidad Católica de Valparaíso (ed.): *El Grial*. Op. cit. (26.10.2016).

2. Kategorisierung der Korpora

Die farbige Anzeige wird als *manchado* bezeichnet und gibt dem *tool* damit seinen Namen. Mittels dieser Option kann nicht nur der typische Kontext der gesuchten Sequenz ermittelt werden, sondern auch die Art und Weise, wie diese sich in einen Text eingliedert. Zudem kann ermittelt werden, in welchen Textabschnitten die Sequenz besonders häufig vorkommt, was im Korpus als ‚linguistische Dichte‘ (*densidad lingüística*) bezeichnet wird.⁴⁸¹

Bei der letzten Option, *Visualización Output*, kann sich der Nutzer Texte vertikal anzeigen lassen, das heißt, die einzelnen Wortformen werden untereinander aufgereiht und horizontal mit Angaben beispielsweise zu deren Position und Funktion im Satz, zur Wortform und zum zugehörigen Lemma versehen:

id_corpus	id_docum	posicion	forma	lema	funcdep	pointer	sst
3315051	9	1	Intramuros	intramuros			@ADVL ADV
3315052	9	2	((
3315053	9	3	Esta	este	det>4		@PREMOD PRON Dem FEM SG
3315054	9	4	noche	noche	tmp>5		@NH N FEM SG
3315055	9	5	estoy	estar	main>0		@MAIN V IND PRES SG P1
3315056	9	6	solo	solo	comp>5		@NH A MSC SG
3315057	9	7))			

Abbildung 53: *El Grial* – Ausschnitt eines Texts in der Option *Visualización Output*.⁴⁸²

Wie man an den verschiedenen Einstellungen sehen kann, bietet dieses Korpus viele Möglichkeiten sowie genaue Analysen der dargebotenen Texte und ist damit ein vielseitig nutzbares Korpus, welches frei und kostenlos im Internet zugänglich ist (<http://www.elgrial.cl>).

2.5.2. Syntaktische Annotation

Eine weitere sehr häufig verwendete Form der Annotation findet auf der syntaktischen Ebene statt. Die Sätze werden hier in einer graphischen Struktur, beispielsweise als Baumdiagramm, dargestellt. Als Beispiele für spanische Korpora mit syntaktischer Annotation werden im Folgenden die Korpora *Base de Datos Sintácticos del Español Actual*, *Europarl-es* und *AnCora* vorgestellt.

⁴⁸¹ Vgl. Pontificia Universidad Católica de Valparaíso (ed.): *El Grial. El Manchador de Textos*. Online Publikation: <http://elgrial.cl/grial/#el-manchador> (06.03.2017).

⁴⁸² Pontificia Universidad Católica de Valparaíso (ed.): *El Grial*. Op. cit. (26.10.2016).

2.5.2.1. BDS – Base de Datos Sintácticos del Español Actual

Die *Base de Datos Sintácticos del Español Actual* der *Universidad de Santiago de Compostela* ist ein manuell annotiertes, syntaktisches Korpus aus ungefähr 160.000 Sätzen, die den aktuellen Teil des *Archivo de Textos Hispánicos de la Universidad de Santiago* (ARTHUS) darstellen.⁴⁸³ Es umfasst 1.450.000 Formen in 34 Texten, die aus den Bereichen Narrativik, Essay, Journalismus und Theater oder aus mündlichen Quellen stammen.⁴⁸⁴ Das Korpus ist im Internet frei verfügbar (<http://www.bds.usc.es>) und es kann darin vor allem nach Verben in deren spezieller Umgebung gesucht werden.

In diesem Korpus können beispielsweise Verben mit bestimmten Objekten oder in gewissen Schemata ermittelt werden. Zudem kann man sich Beispiele für die gesuchten Verben anzeigen lassen oder gezielt nach Verben, die mit speziellen Präpositionen vorkommen, suchen.

Verbos documentados con la preposición "hasta"

Verbo	Frecuencia
ACERCAR	4
ACOMPañAR	13
ACUDIR	1
ANDAR	1
ARRASTRAR	8
ARRIBAR	1
AVANZAR	4
BAJAR	7
CAER	1
CAMINAR	1
COLAR	1
COLGAR	1
CONducIR	7
CONTAR	1
CONTINUAR	1
CORRER	2
DESCENDER	2
DESPLAZAR	1
DESVIAR	1

Abbildung 54: BDS – Suche nach Verben, die mit der Präposition *hasta* verwendet werden.⁴⁸⁵

Bei der Suche nach Verben, die mit der Präposition *hasta* verwendet werden, liefert das Korpus 39 Verben in 226 Vorkommen. In der dazugehörigen Liste werden die Verben alphabetisch aufgelistet und die jeweilige Anzahl daneben angegeben.

⁴⁸³ Vgl. Universidad de Santiago de Compostela (ed.2001): *Base de datos sintácticos del español actual*. Online Publikation: <http://www.bds.usc.es> (25.10.2016).

⁴⁸⁴ Vgl. Universidad de Santiago de Compostela (ed.2001): *Arthus. Composición del corpus*. Online Publikation: <http://www.bds.usc.es/corpus.html> (23.11.2016).

⁴⁸⁵ Universidad de Santiago de Compostela (ed.2001): *Arthus. Buscar verbos con ,hasta'*. Online Publikation: <http://www.bds.usc.es/consultas/preposicion.html> (13.03.2017).

2.5.2.2. Europarl-es

Ein weiteres syntaktisch annotiertes Korpus ist das *Europarl-es*, welches vom Forschungs- und Entwicklungsprojekt *Visual Interactive Syntax Learning* (VISL) des *Institute of Language and Communication* (ISK) der *Syddansk Universitet* (SDU) entwickelt wurde und in der Korpusammlung *Corpuseye* bereitgestellt wird. Es ist im Internet frei zugänglich (<http://corp.hum.sdu.dk/cqp.es.html>) und umfasst ungefähr 27,2 Mio. Lexeme.⁴⁸⁶ Das *Europarl* ist ein Korpus aus Texten parlamentarischer Debatten in unterschiedlichen Sprachen, und es bietet morphosyntaktische und syntaktische Annotationen.⁴⁸⁷

Aufgrund der vielseitigen Annotationen können in diesem Korpus genau spezifizierte Suchen im Bereich der Syntax und der Morphosyntax durchgeführt werden. Bei der Ansicht der Konkordanzen werden das jeweilige Lemma, die Wortart und die Funktion im Satz angezeigt, wenn man mit der Maus über ein Lexem fährt:

The screenshot shows the Corpuseye web interface. At the top, there's a search bar with the text "Searched for: [word='ojo' %c]" and "In corpus: HIS_EUROPAL". Below this, it says "Found 106 results (106)". On the left, there's a sidebar with filters like "sort", "freq", "rel", and "By". The main area displays a list of concordance results. Each result includes a snippet of text and a grammatical analysis of the word "ojo". For example, one result shows "técnicos . Esta es una política trazada a ojo de buen cubero y que se basa única en dirección a este objetivo de abril los ojo y las mentes de nuestros jóvenes y entre el original y la copia . Se necesita un ojo muy adiestrado .". The grammatical analysis for "ojo" in this context is "N MS @P<".

Abbildung 55: *Europarl-es* – Konkordanzen des Substantivs *ojo* mit zusätzlicher Information zu Wortart und Funktion im Satz.⁴⁸⁸

Des Weiteren erhält man durch einen Klick auf den Link „Info“ am linken Rand jeder Konkordanz eine vollständige grammatikalische Analyse des Textausschnitts sowie Informationen zur Textquelle:

⁴⁸⁶ Vgl. VISL (ed.): *Europarl-es*. Online Publikation: <http://corp.hum.sdu.dk/cqp.es.html> (05.12.2016).

⁴⁸⁷ Vgl. VISL (ed.): *Corpus Search Interface*. Online Publikation: http://visl.sdu.dk/corpus_linguistics.html#Spanish_corpora (05.11.2016).

⁴⁸⁸ VISL (ed.): *Europarl-es*. *Buscar „ojo“*. Online Publikation: http://corp.hum.sdu.dk/cgi-bin/cqp.cgi?utf8=√&lang=es&corpus=HIS_EUROPAL&search=ojo&type=basic_search&case=insensitive&Action=Search&password= (13.03.2017).

Mientras	no	lo	haga	y	sea	tuerta	de	un	ojo	,	no	será	un	medio	independiente	.
mientras	no	lo	hacer	y	ser	tuerta	de	un	ojo	no	ser	ser	un	medio	independiente	
ADV	ADV	PERS	V	KC	V	N	PRP	DET	N	ADV	V	DET	N	ADJ		
		M 3S ACC	PR 13S SUBJ		PR 13S SUBJ	M S		M S M S			FUT 3S IND	M S M S	MF S			
@ADVL>	@ADVL>	@ACC>	@FMV	@CO	@FMV	@<SC	@N<	@>N	@P<	@ADVL>	@FMV	@>N	@<SC	@N<		

Abbildung 56: *Europarl-es* – Analyse eines Textausschnitts.⁴⁸⁹

Wie bei anderen Korpora können die Ergebnisse nach dem Umfeld rechts oder links des Suchbegriffs sortiert werden, und der Nutzer kann sich die jeweilige Häufigkeit der Lexeme in diesem Umfeld anzeigen lassen. Zudem kann die Suche mittels der *Refined Search* präzisiert werden: Es können die Grundform, die Wortart sowie morphosyntaktische und syntaktische Informationen zur weiteren Einschränkung der Ergebnisse angegeben werden.

Daneben besteht die Möglichkeit, mehrere *Executive Search Windows* zu öffnen, um mehrgliedrige Suchausdrücke, beispielsweise das Verb *trabajar* mit darauffolgender Präposition, zu suchen. Es können auch mehrere Wortarten angeklickt werden, wodurch beispielsweise nach einem Adjektiv *oder* einem Adverb gesucht werden kann. In der oberen Leiste des *Executive Search Window* kann eingegeben werden, welche Anzahl gesucht wird, und es können Angaben zur Beachtung von Groß- und Kleinschreibung sowie zu diakritischen Zeichen gemacht werden. Die Auswahl *Neg* steht für die Verneinung dieser Kategorie, so zum Beispiel, wenn man alle Wortarten (außer einer Präposition) suchen möchte. In allen Korpora des *Europarl* kann auch mit regulären Ausdrücken und *Wildcards* gesucht werden.

Auf der Seite von *Corpuseye* sind neben dem Korpus *Europarl-es* noch vier weitere spanische Korpora frei zugänglich, unter anderem das Korpus *Internet with semantic roles*, welches im Bereich ‚Semantische Annotation‘ in Kapitel 2.5.3.2. vorgestellt wird. Die Seite bietet zudem eine *Guided Tour*, in der die wichtigsten Funktionen der Korpora beschrieben werden. Dieser Online-Rundgang ist sehr hilfreich und wurde auch zur Erstellung dieser Beschreibung der Funktionsweise des *Europarl-es* genutzt.⁴⁹⁰

2.5.2.3. AnCora

Das Korpus AnCora ist ein spanisches und katalanisches Korpus mit verschiedenen Ebenen der Annotation. Es ist morphosyntaktisch, syntaktisch und semantisch annotiert. Dabei werden unter anderem Lemmata und Wortarten sowie Satzglieder,

⁴⁸⁹ VISL (ed.): *Europarl-es. Buscar ,ojo‘*. Op. cit. (13.03.2017).

⁴⁹⁰ Vgl. VISL (ed.): *Guided Tour*. Online Publikation: <https://beta.visl.sdu.dk/guide/> (05.12.2016).

2.5.3. Semantische Annotation

Eine weitere Annotationsmethode ist die semantische Annotation. Diese kann entweder auf der Lexemebene stattfinden oder in Form von semantischen *frames* darüber hinausgehen. Die beiden Korpora *SenSem* und *Internet with semantic roles*, welche in diesem Kapitel vorgestellt werden, sowie das im obigen Abschnitt bereits vorgestellte AnCora (siehe Kapitel 2.5.2.3.) sind semantisch annotiert.

2.5.3.1. *SenSem (Español)*

Beim Korpus *SenSem* handelt es sich um ein weiteres, sehr interessantes Spezialkorpus der *Grup de Recerca Interuniversitari en Aplicacions Lingüístiques* (GRIAL), des bereits vorgestellten Arbeitskreises der vier Universitäten Kataloniens, welcher auch das *Corpus Trilingüe Paralelo Grial* (siehe Kapitel 2.3.1.) entwickelt hat.⁴⁹⁴ Es ist im Internet kostenlos und ohne jegliche Anmeldung zugänglich (<http://grial.uab.es/sensem/corpus?idioma=es>).

Die Besonderheit des Korpus *SenSem* besteht in der Miteinbeziehung der Semantik in das Feld der Korpuslinguistik. In diesem Korpus kann nach spanischen und katalanischen Verben gesucht werden, wobei die gefundenen Konkordanzen den unterschiedlichen Bedeutungen der Verben zugeteilt werden. Die Texte im Korpus *SenSem Español* sind vor allem dem journalistischen Bereich, teilweise aber auch der Literatur entnommen. Es umfasst ungefähr eine Mio. Lexeme (30.000 Sätze), wobei diese nach den 250 häufigsten Verben des Spanischen ausgewählt wurden. Die Sätze wurden manuell syntaktisch-semantisch annotiert (im Hinblick auf semantische Rollen, syntaktische Funktionen, syntagmatische Kategorien, Konstruktionen, Aspekt, Modus und Polarität). Die Suchmaske lässt viele Einstellungen zu und es kann nach Verben und/oder linguistischen Phänomenen gesucht werden. So können im Korpus *SenSem* Beispiele für das Verb *contar* in acht Bedeutungen ermittelt werden. Die Verben und deren jeweilige Bedeutungen können in der Suchmaske ausgewählt werden, als Ergebnis werden alle dafür infrage kommenden Sätze aufgelistet, und zusätzlich können die Annotationen der Sätze angezeigt werden.⁴⁹⁵

⁴⁹⁴ Vgl. Grup de Recerca Interuniversitari en Aplicacions Lingüístiques (ed.): *Inicio*. Op. cit. (28.10.2016).

⁴⁹⁵ Vgl. Grup de Recerca Interuniversitari en Aplicacions Lingüístiques (ed.): *Recursos. Corpus*. Op. cit. (21.12.2016).

2. Kategorisierung der Korpora

contar (1 / 8 sentidos)

contar (8): Ser significativo o relevante.

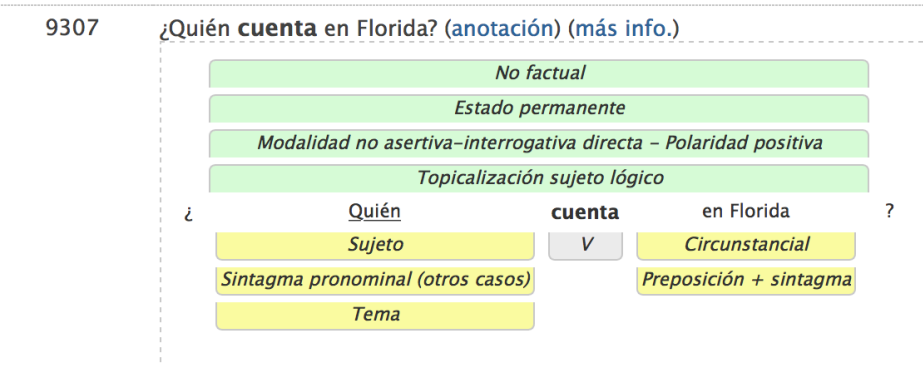


Abbildung 58: *SenSem* – Suche nach dem Verb *contar* in der Bedeutung *ser significativo o relevante*.⁴⁹⁶

Zudem gibt es das Korpus *SenSem Catalán*, bestehend aus 600.000 Lexemen in 20.000 Sätzen. Beide Korpora sind Parallelkorpora, denn das *SenSem Catalán* wurde aus den Übersetzungen der Zeitungstexte des *SenSem Español* erstellt, doch es kann jeweils nur das Korpus einer Sprache angezeigt werden.⁴⁹⁷

2.5.3.2. Internet with semantic roles (VISL)

Neben dem *Europarl-es* (siehe Kapitel 2.5.2.2.) findet man auf der Seite *Corpuseye* des Projekts *Visual Interactive Syntax Learning* (VISL) auch das Korpus *Internet with semantic roles*. Wie der Name bereits verrät, ist dieses Korpus semantisch getaggt und lässt eine Suche nach semantischen Rollen zu. Es umfasst 11,2 Mio. Lexeme und ist im Internet frei zugänglich (<http://corp.hum.sdu.dk/cqp.es.html>).⁴⁹⁸ Aufgrund der Tatsache, dass es sich bei *Corpuseye* um eine Art Korpusammlung handelt, ist die Suchmaske bei allen Korpora gleich aufgebaut, unabhängig davon, ob man nur in einem Korpus oder gleichzeitig in allen Korpora einer Sprache sucht. Die *Refined Search* bietet in diesem Korpus allerdings zusätzlich die Möglichkeit nach semantischen Rollen zu suchen. Es kann beispielsweise nach dem Substantiv *profesor* als *agent*, *speaker* oder *patient* gesucht werden.

⁴⁹⁶ Grup de Recerca Interuniversitari en Aplicacions Lingüístiques (ed.): *Corpus SenSem. Búsqueda*. Online Publikation: <http://grial.uab.es/sensem/corpus?idioma=es> (13.03.2017).

⁴⁹⁷ Vgl. Grup de Recerca Interuniversitari en Aplicacions Lingüístiques (ed.): *Recursos. Corpus*. Op. cit. (21.12.2016).

⁴⁹⁸ Vgl. VISL (ed.): *Internet with semantic roles*. Online Publikation: <http://corp.hum.sdu.dk/cqp.es.html> (05.12.2016).

hijo Carlos prestigioso **profesor** de Historia_Económica y de sus dos nietas Ana_Esperanza
 ✎ Y eso que el **profesor** nos llevaba a correr por los alrededores de los Agustinos
 o porque digamos que el **profesor** suele explicar más sobre el conocimiento y en_cambio
 ✎ Un **profesor** universitario retó a sus alumnos con esta pregunta
 Lo digo porque si algún **profesor** ha dicho eso que apuntas en tu post en clase el principal
 ✎ A mi entender el **profesor** Enguita no sólo no parece estar alejado de las escuelas

Abbildung 59: *Internet with semantic roles* – Ergebnisse der Suche nach dem Substantiv *profesor* als *agent*.⁴⁹⁹

✎ Como nos señala el **profesor** Csikszentmihalyi cuando la energía psic
 a repetir y tomar como bien lo que el **profesor** dice éste es el criterio que se toma a
 o cuando vi el plan contable y que el **profesor** dijo a toda_la clase
 pidió informes acerca_de su alumno el **profesor** respondió
 hace años también se refirió a esto un **profesor** mío en el instituto a que al_contrario_
 Abbildung 60: *Internet with semantic roles* – Ergebnisse der Suche nach dem Substantiv *profesor* als *speaker*.⁵⁰⁰

Todo lo que yo deseaba es que el **profesor** Neira mejorará y al_fin ha despertado de su coma
 cuestión muy diferente es si el **profesor** debe convertirse se o no en autor de estos objet
 realista que el hecho de que el **profesor** al_final se convierta en un informático eso es m
 ✎ El **profesor** Paulo_Sergio_Pinheiro experto independiente fue
 de estos pequeños conflictos este **profesor** se ha limitado a observar los comportamientos y
 Abbildung 61: *Internet with semantic roles* – Ergebnisse der Suche nach dem Substantiv *profesor* als *patient*.⁵⁰¹

Wie man an den drei Abbildungen erkennt, erhält der Nutzer je nach der
 ausgewählten semantischen Rolle diejenigen Sätze, in denen die jeweilige Rolle erfüllt
 ist.

2.6. Größe

Ein weiteres Kriterium der Korpus typologie ist die Größe. Während das erste
 maschinenlesbare Korpus, das *Brown Corpus*, lediglich 1 Mio. Wortformen umfasste,
 sind derzeit Referenzkorpora mit 100 Milliarden Lexemen verfügbar. Man darf sich von
 der Größe eines Korpus aber nicht täuschen lassen, denn je größer ein Korpus ist, desto

⁴⁹⁹ VISL (ed.): *Internet with semantic roles*. Buscar ‚profesor‘ & ‚agent‘. Online Publikation:
http://corp.hum.sdu.dk/cgi-bin/cqp.cgi?corpus=HIS_WEB_ROLE&lang=es&utf8=✓&type=refine_search&search=%5Bword%3D%22profesor%22+%26+role%3D%22%28.*+%29%3FAG.*%28+.*%29%3F%22+%25c%5D
 (13.03.2017).

⁵⁰⁰ VISL (ed.): *Internet with semantic roles*. Buscar ‚profesor‘ & ‚speaker‘. Online Publikation:
http://corp.hum.sdu.dk/cgi-bin/cqp.cgi?corpus=HIS_WEB_ROLE&lang=es&utf8=✓&type=refine_search&search=%5Bword%3D%22profesor%22+%26+role%3D%22%28.*+%29%3FSP%28+.*%29%3F%22+%25c%5D
 (13.03.2017).

⁵⁰¹ VISL (ed.): *Internet with semantic roles*. Buscar ‚profesor‘ & ‚patient‘. Online Publikation:
http://corp.hum.sdu.dk/cgi-bin/cqp.cgi?corpus=HIS_WEB_ROLE&lang=es&utf8=✓&type=refine_search&search=%5Bword%3D%22profesor%22+%26+role%3D%22%28.*+%29%3FPAT%28+.*%29%3F%22+%25c%5D
 (13.03.2017).

2. Kategorisierung der Korpora

weniger ist es für gewöhnlich annotiert. Design und Größe des Korpus hängen letztendlich auch von der Fragestellung ab. Für einige Projekte sind möglichst große Korpora unabdingbar, für andere ist ein kleines, gut annotiertes Spezialkorpus weitaus gewinnbringender.⁵⁰²

Für die spanische Sprache gibt es einige umfassende Referenzkorpora, wie das CREA (mehr als 154 Mio. Wortformen; siehe Kapitel 2.1.1.1.)⁵⁰³ und das CORPES (mehr als 225 Mio. Wortformen im Juni 2016, letztendlich soll es im Jahr 2018 über 400 Mio. Lexeme umfassen; siehe Kapitel 2.1.1.2.)⁵⁰⁴. Außerdem gibt es zwei große historische Korpora, das CORDE der *Real Academia Española* (siehe Kapitel 2.2.1.1.) mit seinen 250 Mio. Eintragungen (April 2015)⁵⁰⁵ und das CdE:Orig von Mark Davies (siehe Kapitel 2.2.1.3.) mit 100 Mio. Lexemen.⁵⁰⁶ Neben diesen umfangreichen Korpora gibt es für das Spanische noch eine Vielzahl kleinerer Spezialkorpora:

- Das morphosyntaktisch annotierte Korpus *El Grial* der *Pontificia Universidad Católica de Valparaíso* umfasst 35 Dokumente mit insgesamt 5,4 Mio. Lexemen⁵⁰⁷ wohingegen das hinsichtlich der Annotation vergleichbare *Corpus Tècnic del IULA* der *Universitat Pompeu Fabra* über 1813 spanische Texte mit mehr als 33 Mio. Lexemen verfügt (siehe Kapitel 2.5.1.).⁵⁰⁸
- Das syntaktisch annotierte Korpus BDS der *Universidad de Santiago de Compostela* umfasst 1.450.000 Formen in 160.000 Sätzen⁵⁰⁹ und das ebenfalls syntaktisch annotierte Korpus *Europarl-es* ganze 27,2 Mio. Lexeme.⁵¹⁰ Das morphosyntaktisch, syntaktisch und semantisch annotierte Korpus AnCora zählt für jede Sprache jeweils 500.000 Lexeme (siehe Kapitel 2.5.2.).⁵¹¹

⁵⁰² Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 105.

⁵⁰³ Vgl. Real Academia Española (ed.): *CREA. Estadísticas*. Op. cit. (14.05.2011).

⁵⁰⁴ Vgl. Real Academia Española (ed.): *CORPES*. Op. cit. (24.10.2016).

⁵⁰⁵ Vgl. Real Academia Española (ed.): *Banco de datos. CORDE*. Op. cit. (19.05.2011).

⁵⁰⁶ Vgl. Davies: *Corpus del Español. Corpora*. Op. cit. (21.11.2016).

⁵⁰⁷ Vgl. Pontificia Universidad Católica de Valparaíso (ed.): *El Grial. Interfaz de etiquetaje e interrogación de corpus textuales*. Op. cit. (26.10.2016).

⁵⁰⁸ Vgl. Institut Universitari de Lingüística Aplicada (ed.): *BwanaNet. Programa de explotación del Corpus Técnico de IULA*. Op. cit. (26.10.2016).

⁵⁰⁹ Vgl. Universidad de Santiago de Compostela (ed.2001): *Arthus. Composición del corpus*. Op. cit. (23.11.2016).

⁵¹⁰ Vgl. VISL (ed.): *Europarl-es*. Op. cit. (05.12.2016).

⁵¹¹ Vgl. Grupo de investigación CLiC (ed.): *AnCora*. Op. cit. (12.01.2017).

- Die semantisch annotierten Korpora *SenSem* und *Internet with semantic roles* enthalten eine Mio.⁵¹² beziehungsweise 11,2 Mio. Lexeme⁵¹³ (siehe Kapitel 2.5.3.).
- Bei den Korpora geschriebener Sprache ist es relativ leicht, die Größe anhand der Lexeme anzugeben, so umfasst das CEA 540 Mio.⁵¹⁴ und das *Hemero* 660 Mio. Lexeme⁵¹⁵ (siehe Kapitel 2.4.1.). Bei den Korpora gesprochener Sprache ist die Angabe der Lexemanzahl schwerer. So ist zwar bekannt, dass der spanische Teil des C-ORAL-ROM 300.000⁵¹⁶, das Korpus CORLEC 1,1 Mio.⁵¹⁷ und das C-Or-DiAL knapp 119.000 Lexeme⁵¹⁸ umfasst, doch für das *Macrocorpus* kann lediglich angegeben werden, dass es sich um 84 Stunden gesprochener spanischer Sprache handelt (siehe Kapitel 2.4.2.).

Neben all diesen Korpora der spanischen Sprache gibt es drei Korpora, die speziell im vorliegenden Unterkapitel beschrieben werden sollen. Erstens das CdE:Nuevo von Mark Davies, welches erst seit dem Sommer 2016 verfügbar ist und mit seinen 2 Milliarden Lexemen die bisher verfügbaren Referenz- und opportunistischen Korpora in Bezug auf die Größe um einiges übertrifft. Zweitens das ebenfalls von Davies entwickelte Korpus *Google Books: Spanish*, welches nicht eigens als Korpus zusammengestellt wurde, sondern auf Daten aus *Google Books* zurückgreift. Es verfügt zwar nicht über alle Suchoptionen der anderen Korpora von Davies, setzt dafür aber mit seinen 45 Milliarden Lexemen neue Maßstäbe.⁵¹⁹ Und drittens das ebenfalls von Davies erstellte und erst im August 2018 veröffentlichte *NOW Corpus: Spanish (News on the Web)*, welches in seiner derzeitigen Version 5,5 Milliarden Lexeme aus aktuellen, online verfügbaren Magazinen und Zeitschriften umfasst und fortlaufend erweitert werden soll.⁵²⁰

⁵¹² Vgl. Pontificia Universidad Católica de Valparaíso (ed.): *El Grial. Interfaz de etiquetaje e interrogación de corpus textuales*. Op. cit. (26.10.2016).

⁵¹³ Vgl. VISL (ed.): *Internet with semantic roles*. Op. cit. (05.12.2016).

⁵¹⁴ Vgl. Subirats/ Ortega: *Corpus del Español Actual (CEA)*. Op. cit. (17.03.2017).

⁵¹⁵ Vgl. Molino de Ideas (ed.): *Hemero.es - El corpus molinero*. Op. cit. (14.03.2017).

⁵¹⁶ Vgl. Directorate General Information Society (ed.): op. cit. (03.12.2011).

⁵¹⁷ Vgl. Laboratorio de Lingüística Informática de la Universidad Autónoma de Madrid (ed.): *Corpus Oral de Referencia de la Lengua Española Contemporánea (CORLEC)*. Op. cit. (24.11.2016).

⁵¹⁸ Vgl. Laboratorio de Lingüística Italiana (ed.): *Corpus Didáctico Anotado Lingüísticamente (C-Or-DiAL)*. Op. cit. (19.03.2017).

⁵¹⁹ Vgl. Davies, Mark: *El corpus del español. Corpora*. Online Publikation: <http://www.corpusdelespanol.org/x.asp?c=4> (20.01.2018).

⁵²⁰ Vgl. Davies, Mark: *El corpus del español. Corpora. NOW*. Online Publikation: <http://www.corpusdelespanol.org/xs.asp?c=4> (06.10.2018).

2.6.1. CdE:Nuevo – *Corpus del Español (Web/Dialectos)*

Wie bereits erwähnt, besteht das CdE von Mark Davies seit dem Sommer 2016 aus zwei verschiedenen Teilen.⁵²¹ Das ursprüngliche CdE, welches nun vom Autor als *Corpus del Español (Género/Histórico)* oder CdE:Orig bezeichnet wird, wurde bereits im Rahmen der historischen Korpora genauer beschrieben (siehe ‚CdE:Orig (Gesamt)‘ in Kapitel 2.2.1.3.). Zudem stellt der darin enthaltene Teil des aktuellen Spanisch ein Referenzkorpus dar (siehe Kapitel 2.1.1.3.). Seit 2016 gibt es zusätzlich ein neues Korpus, das *Corpus del Español (Web/Dialectos)* oder CdE:Nuevo, welches knapp zwei Milliarden Lexeme aus Texten aus 21 spanisch-sprachigen Ländern umfasst und damit hundertmal so groß ist wie der aktuelle Teil des CdE:Orig mit 20 Mio. Lexemen. Es handelt sich um Texte von zwei Mio. aktuellen Internetseiten, welche Ende 2015 gesammelt wurden. Das neue Korpus und das Update des ursprünglichen Korpus wurden vom *US National Endowment for the Humanities* finanziert.⁵²²

Es ist allerdings unklar, ob es sich beim CdE:Nuevo um ein Referenzkorpus oder ein opportunistisches Korpus handelt (für eine Gegenüberstellung der beiden Korpusarten siehe Kapitel 2.1.). Der Ersteller, Mark Davies, äußert sich in der Beschreibung des Korpus nicht dazu, doch im Hinblick auf die Größe könnte es mit 2 Milliarden Lexemen in beide Kategorien eingeordnet werden. Auf den Internetseiten, die das Korpus beschreiben, findet sich ein Vergleich zwischen dem CdE:Nuevo und dem CORPES der *Real Academia Española*. In diesem Vergleich mit dem Referenzkorpus CORPES gibt Davies an, sein Korpus verfüge über mehr Formen (knapp 2 Milliarden im Vergleich zu 225 Mio. Lexemen), Texte aus ebenso vielen Ländern (21 Länder in beiden Korpora, wobei es im CORPES zusätzlich Texte aus Äquatorial-Guinea und den Philippinen gibt) und ein besseres Verhältnis zwischen Texten aus Lateinamerika und Spanien (78% zu 22% im Vergleich zu 65% und 35%⁵²³). Darüber hinaus kritisiert Davies die Erstellung und Zusammensetzung des CORPES auf seiner Internetseite, beispielsweise wenn er anführt, dass die 225 Mio. Formen des CORPES eigentlich nur 175 Mio. sind. Zu dieser Erkenntnis sei er aufgrund eines Vergleichs der Suchergebnisse

⁵²¹ Vgl. Davies: *Corpus del Español. Corpora*. Op. cit. (21.11.2016).

⁵²² Vgl. Davies, Mark: *Corpus del Español. Virtual Corpus*. Online Publikation: <http://www.corpusdelespanol.org/web-dial/> (21.11.2016).

⁵²³ Anmerkung: Davies gibt den Prozentsatz 65% zu 35% an, wobei auf den Seiten des CORPES ein Prozentsatz von 70% aus Lateinamerika und 30% aus Spanien festgehalten wird.

für frequente Lexeme wie *de* oder *en* in beiden Korpora gelangt. Zudem seien die zusätzlichen Texte aus Äquatorial-Guinea und den Philippinen, bei denen es sich lediglich um 640.000 beziehungsweise 100.000 Lexeme handelt, zu wenige, um diese Länder mitzuzählen. Des Weiteren begründet er sein Verhältnis zwischen den Texten aus Spanien und Lateinamerika damit, dass nur 10% der spanischsprachigen Bevölkerung in Spanien lebt und gibt an, die Favorisierung Spaniens bei der geographischen Aufteilung der Textanzahl im CORPES sei der *Real Academia Española* geschuldet.⁵²⁴ Ungeachtet der Kritik, die Davies am CORPES äußert, könnte das CdE:Nuevo aufgrund der genannten Informationen durchaus als Referenzkorpus gesehen werden, doch bezüglich der Textauswahl ist die Einordnung nicht ganz klar.

Davies betont, dass es sich um sehr aktuelle Texte handelt, da alle Texte des Korpus aus den Jahren 2013 und 2014 stammen.⁵²⁵ Somit beschränkt sich die Textauswahl des CdE:Nuevo allerdings auf Daten aus zwei Jahren, während die Texte des CREA 25 Jahre (beziehungsweise 30 Jahre in der Version von 2015)⁵²⁶ und diejenigen des CORPES 12 Jahre abdecken.⁵²⁷ Zudem kann sich der Nutzer zwar die Metadaten der Webseiten (Land, Gattung, Internetadresse, Titel sowie Anzahl der Lexeme) als Liste herunterladen, es finden sich aber keine Informationen zur Kategorisierung der Texte oder zur Begründung der Textauswahl, abgesehen von der bereits erwähnten Aktualität derselben. Es wird lediglich angegeben, dass es die beiden Kategorien ‚allgemeine Texte‘ und ‚Blogs‘ gibt, wobei sich auch unter den allgemeinen Texten Blogs befinden können.⁵²⁸ Aus diesem Grund kann nicht genau gesagt werden, dass auch auf Literatur oder mündliche Texte verschiedener Gattungen geachtet wurde, was eine Voraussetzung für die Erstellung eines Referenzkorpus wäre. Diese fehlenden Angaben zur Textauswahl und die ungenaue Beschreibung des Korpus lassen somit keine Rückschlüsse auf die Repräsentativität und Ausgeglichenheit des Korpus zu.

Im Internetauftritt des CdE:Nuevo wird es als größtes spanisches Korpus mit vollständiger, sorgfältiger Lemmatisierung und *POS-tagging* beworben, wobei der eben

⁵²⁴ Vgl. Davies, Mark: *Corpus del Español – Web/Dialectos*. Online Publikation: <http://www.corpusdelespanol.org/web-dial/> (16.12.2016).

⁵²⁵ Vgl. Davies: *Corpus del Español – Web/Dialectos*. Op. cit. (16.12.2016).

⁵²⁶ Vgl. Real Academia Española (ed.): *CREA. Ayuda*. Op. cit. (14.05.2011).

⁵²⁷ Vgl. Real Academia Española (ed.): *Banco de datos. CORPES*. Op. cit. (24.10.2016).

⁵²⁸ Vgl. Davies: *Corpus del Español – Web/Dialectos*. Op. cit. (16.12.2016).

dargestellte Vergleich mit dem CORPES die Vorteile des CdE:Nuevo hervorheben soll.⁵²⁹ Davies beschreibt das CORPES zwar als sehr gutes Korpus, vor allem im Hinblick auf die Textauswahl und die Kategorisierung der Texte, er kritisiert jedoch die Korpusgröße sowie die seiner Meinung nach ‚rudimentäre‘ Benutzeroberfläche. Für ihn kann das lobenswerte Korpus aufgrund des ärmlichen Interfaces nicht ausreichend genutzt werden.⁵³⁰ Auf den Seiten seiner Korpora wird schnell klar, dass für Davies insbesondere die Größe der Korpora ausschlaggebend ist. Weder die Textzusammenstellung noch die Genauigkeit der Annotation sind für ihn von derartiger Relevanz. Fraglich ist jedoch, ob seine Ansicht von den Nutzern geteilt wird, da seine Korpora zwar sehr benutzerfreundlich sind und viel zu bieten haben, die Zuverlässigkeit der Annotation sowie die Textauswahl, und damit die Repräsentation der dargestellten Sprache, jedoch durchaus eine wichtige Rolle spielen. Aus eigener Erfahrung kann gesagt werden, dass eben diese Kriterien durchaus wichtig sind hinsichtlich der Auswahl eines Korpus für eigene Untersuchungen. Die Größe eines Korpus stellt in diesem Fall nicht das bedeutendste Kriterium dar, da ein großes, dafür aber fehlerbehaftetes oder schlecht annotiertes Korpus nicht gewinnbringend genutzt werden kann.

Im Gegensatz zum opportunistischen Korpus *Spa_news_2011_3M* des *FCS Aggregator*, welches zwar lemmatisiert ist, aber keinerlei Annotation bietet, und schlichtweg eine Vereinigung verschiedener Korpora darstellt, ist das CdE:Nuevo allerdings sehr geplant und gut annotiert, was die Einordnung in die Korpustypologie zusätzlich erschwert.

Aufgrund der Tatsache, dass es sich um ein äußerst umfassendes Korpus mit Texten aus derart vielen spanischsprachigen Ländern und einer wohlüberlegten, an den tatsächlichen Bevölkerungszahlen orientierten Aufteilung in spanische und lateinamerikanische Texte handelt, würde eine Einordnung als Referenzkorpus naheliegen, doch die mangelnde Information zur Gattung der Texte, deren Verhältnis zueinander und zur Repräsentativität des Korpus führen tendenziell zur Kategorisierung desselben als opportunistisches Korpus. Man könnte es also als Mischung aus Referenzkorpus und opportunistischem Korpus sehen. Aufgrund seiner Größe, der

⁵²⁹ Vgl. Davies, Mark: *Full-text corpus data from corpus.byu.edu*. Online Publikation: <http://corpus.byu.edu/full-text/spanish.asp> (13.12.2016).

⁵³⁰ Vgl. Davies, Mark: *Corpus del Español – Web/Dialectos. Comparación*. Online Publikation: http://www.corpusdelespanol.org/web-dial/help/compare_s.asp (16.01.2017).

geographischen Vielfalt und der zahlreichen Nutzungsmöglichkeiten ist es aber auf jeden Fall ein für den Nutzer sehr wertvolles Korpus.

Im CdE:Nuevo sind, wie beim CdE:Orig, zunächst 10 bis 15 Eingaben ohne vorherige Anmeldung möglich. Danach muss der Benutzer sich registrieren und ein Profil erstellen. Die Benutzung des Korpus ist zwar kostenlos, die Registrierung ist allerdings unumgänglich (<https://www.corpusdelespanol.org/web-dial/>). Bei der Erstellung des Profils muss man seinen Namen angeben und einige Fragen zur eigenen Person sowie zum Zweck der Nutzung des Korpus beantworten. Außerdem kann das CdE:Nuevo gegen eine Gebühr von 245-795 \$ (je nach Nutzungsart) für die offline Nutzung heruntergeladen werden.⁵³¹ Wenn sich der Nutzer dazu entschließt, hat er die vollständigen Texte des Korpus dauerhaft vorliegen und kann diese für seine Arbeit verwenden.

Im CdE:Nuevo kann sowohl nach einzelnen Wortformen als auch nach Lemmata, Ausdrücken mit *Wildcards* und ganzen Syntagmen gesucht werden. Zudem erlaubt es die Suche nach Synonymen und Kollokationen, und auch personalisierte Wortlisten können verwendet werden. Die Treffer kann sich der Nutzer als Liste oder als Grafik anzeigen lassen. Des Weiteren können Ergebnisse aus unterschiedlichen Dialekten beziehungsweise Regionen miteinander verglichen werden.⁵³² Als äußerst interessant und bisher einzigartig stellt sich beim CdE die Möglichkeit der Suche nach Synonymen heraus. So kann im CdE:Nuevo mittels der Nutzung eines regulären Ausdrucks nach Synonymen gesucht werden:



Abbildung 62: CdE:Nuevo – Suche nach Synonymen für das Lemma *guapo*.⁵³³

Die Kombination aus Gleichheitszeichen und einem Ausdruck in Großbuchstaben führt zur Anzeige aller erfassten Synonyme des Lemmas *guapo*. Es werden also alle Formen möglicher Adjektive angezeigt, seien sie maskulin, feminin, im Singular oder Plural:

⁵³¹ Vgl. Davies, Mark: *Full-text corpus data from corpus.byu.edu – Purchase data*. Online Publikation: <http://corpus.byu.edu/full-text/purchase.asp> (13.12.2016).

⁵³² Vgl. Davies: *Corpus del Español – Web/Dialectos*. Op. cit. (16.12.2016).

⁵³³ Davies: *Corpus del Español – Web/Dialectos*. Op. cit. (16.12.2016).

2. Kategorisierung der Korpora

CONTEXTO	FREC
GENIAL [S]	77249
BONITO [S]	64577
HERMOSA [S]	61962
HERMOSO [S]	60827
LINDO [S]	52440
BELLA [S]	52363
ATRACTIVO [S]	46928
BELLO [S]	43925
BONITA [S]	41395
LINDA [S]	41158
VALIENTE [S]	32039
ELEGANTE [S]	25858

Abbildung 63: CdE:Nuevo – Beginn der Übersicht der Synonyme für das Lemma *guapo*.⁵³⁴

Das CdE:Nuevo ermöglicht jedoch nicht nur die Anzeige der Synonyme als Frequenzliste, sondern auch einen Vergleich zwischen einzelnen Varietäten des Spanischen. Hierfür muss in der Suchmaske lediglich die Option *Secciones* angeklickt werden (siehe Abbildung 62 unten links).

	<input type="checkbox"/>	CONTEXTO	TODOS <input type="checkbox"/>	Argentina <input type="checkbox"/>	Bolivia <input type="checkbox"/>	Chile <input type="checkbox"/>	Colombia <input type="checkbox"/>	Costa Rica <input type="checkbox"/>	Cuba <input type="checkbox"/>	Rep Dom <input type="checkbox"/>	Ecuador <input type="checkbox"/>	España <input type="checkbox"/>
1	<input type="checkbox"/>	GENIAL [S]	77249	6250	503	2599	3049	843	1133	589	952	33700
2	<input type="checkbox"/>	BONITO [S]	64577	1790	548	1919	5106	1259	1141	790	1134	22042
3	<input type="checkbox"/>	HERMOSA [S]	61962	6093	1238	2249	6054	1089	2419	1202	1685	8978
4	<input type="checkbox"/>	HERMOSO [S]	60827	6026	1064	2295	6196	979	1658	974	1625	8982
5	<input type="checkbox"/>	LINDO [S]	52440	10542	690	3260	3738	865	996	628	1322	5135
6	<input type="checkbox"/>	BELLA [S]	52363	3671	1102	1489	4711	768	1623	872	1206	9940
7	<input type="checkbox"/>	ATRACTIVO [S]	46928	3281	791	1633	3631	880	1194	891	1157	12134
8	<input type="checkbox"/>	BELLO [S]	43925	2824	617	2226	5132	684	1421	829	1053	6340
9	<input type="checkbox"/>	BONITA [S]	41395	1393	501	1146	3536	598	866	532	760	12955
10	<input type="checkbox"/>	LINDA [S]	41158	5970	769	2751	3501	747	1033	541	1082	3532
11	<input type="checkbox"/>	VALIENTE [S]	32039	2246	514	807	2298	530	1417	494	803	7774
12	<input type="checkbox"/>	ELEGANTE [S]	25858	1498	331	834	2179	411	974	467	624	7300

Abbildung 64: CdE:Nuevo – Übersicht der Synonyme für das Lemma *guapo* im Vergleich der Varietäten.⁵³⁵

Das Ergebnis ist eine Liste der Synonyme, welche nicht nur nach der gesamten Häufigkeit der einzelnen Ausdrücke, sondern nach den Treffern für jede Varietät aufgeschlüsselt ist. Derartige Vergleiche können bisweilen nützlich sein, um zu überprüfen, in welchen Varietäten bestimmte Ausdrücke besonders häufig verwendet werden und in welchen nicht.

⁵³⁴ Davies: *Corpus del Español – Web/Dialectos*. Op. cit. (16.12.2016).

⁵³⁵ Davies: *Corpus del Español – Web/Dialectos*. Op. cit. (16.12.2016).

Zudem bietet das CdE:Nuevo seit dem Jahr 2017 das Programm *WordAndPhrase*, welches Einträge der 40.000 häufigsten Lexeme des Spanischen enthält. Die Lexeme wurden anhand der Frequenz im Korpus ermittelt, und ihre Einträge enthalten eine Definition, Synonyme, die Häufigkeit nach Genre und nach Region, Kollokationen, Themen (in diesem Fall Lexeme, die auf derselben Internetseite vorkommen) sowie 200 Beispielkonkordanzen. Die Lexeme sind miteinander verlinkt, wodurch der Nutzer von einem Eintrag direkt zum Eintrag eines dort angegebenen Synonyms gelangen kann. Des Weiteren kann auch mit Platzhaltern gesucht werden.⁵³⁶ Der Eintrag des Substantivs *problema* sieht wie folgt aus:

PROBLEMA *n* (RANK 112, FREQ 1325371)

DEFINITIONS (click for more)
problem worry (question)

Corrections / problems ?

SPOK **FIC** **NEWS** **ACAD**

SYNONYMS (click to see word)

acertijo *n* **puzzle** rompecabezas, jeroglífico, crucigrama, solitario, charada, pasatiempo,
 aprieto *n* (**pressure**) conflicto, ahogo, dilema, apuro,
 apuro *n* **predicament** dificultad, trance, brete, aprieto, compromiso,
 detención *n* **detention** atasco, dificultad, retraso, demora, estancamiento, interrupción, tropiezo,
 dificultad *n* **problem** apuro, apremio, contrariedad, inconveniente, engorro, tropiezo, conflicto,
 dilema *n* **quandary** dificultad, conflicto, contrariedad, alternativa, enigma, inconveniente,
 disyuntiva *n* (**choice**) conflicto, dificultad, contradicción, complejidad, dilema,
 duda *n* **question** pregunta, cuestión, incógnita, enigma,
 incógnita *n* **unknown** enigma, misterio, secreto, entresijo, (quisicosa),
 inconveniente *n* **problem** contrariedad, obstáculo, engorro, escollo, atolladero, dificultad,
 obstáculo *n* **hurdle** inconveniente, tropiezo, dificultad, escollo,
 vicisitudes *n* (-----) azar, dilema, fortuna, dificultad, variación, altibajos,

COLLOCATES ☒ new word ☐ with PROBLEMA

(NOUN) solución, salud, resolución, inseguridad, obesidad, adicción, sobrepeso, tiroides, fertilidad, abastecimiento
 (ADJ) grave, serio, estructural, respiratorio, cardíaco, digestivo, cardiovascular, resuelto, renal, cardíaco
 (MISC) resolver, solucionar, enfrentar, causar, plantear, afrontar, abordar, radicar, ocasionar, arreglar

TOPICS (click to see)

síntoma *n* trastorno *n* instalar *v* paciente *n* solucionar *v* tratamiento *n* medicamento *n*
 pareja *n* sexual *j* droga *n* archivo *n* dieta *n* embarazo *n* aprendizaje *n* solución *n*
 médico *n* software *n* escolar *j* terapia *n* alimento *n* alumno *n* resolver *v* enfermedad *n*
 infección *n* bebé *n* docente *n* consumo *n* ambiental *j* grasa *n* violencia *n*

Abbildung 65: *WordAndPhrase* – Eintrag des Substantivs *problema*.⁵³⁷

Unter dem abgebildeten Eintrag sind ein Diagramm mit der Frequenz des Lexems in den einzelnen spanischsprachigen Ländern sowie Beispielkonkordanzen aufgelistet, anhand derer für den Nutzer ersichtlich ist, wie das jeweilige Lexem verwendet wird und in welchem Kontext es vorkommt.

Wird auf das Fragezeichen am Rand der Eingabezeile geklickt, so zeigt das Korpus den Eintrag eines willkürlich ausgewählten Lexems an. Das Programm kann ebenso wie ein Wörterbuch Englisch-Spanisch genutzt werden. Es kann also ein englisches Lexem eingegeben werden, und das Programm zeigt die spanische Bedeutung beziehungsweise die spanischen Bedeutungen an. Das gesamte Programm basiert auf Häufigkeiten, so kann

⁵³⁶ Vgl. Davies, Mark: *WordAndPhrase*. Online Publikation: <https://www.wordandphrase.info/span/> (20.01.2018).

⁵³⁷ Davies: *WordAndPhrase*. Op. cit. (20.01.2018).

unter anderem nach häufig vorkommenden Verben, mittelmäßig häufigen Substantiven oder seltenen Adjektiven gesucht werden.⁵³⁸

Sofern die Realisierung der gegenwärtig laufenden Projekte wie geplant verläuft, soll eventuell auch ein spanisches Wikipedia-Korpus ähnlich dem bereits existierenden englischsprachigen Pendant erstellt werden. Auf diese Weise könnten die Benutzer schnell und einfach ein virtuelles Korpus zu beliebigen Themengebieten erstellen und daraus wichtige Schlagwörter extrahieren.⁵³⁹

2.6.2. Google Books: Spanish

Ein weiteres umfassendes opportunistisches Korpus ist das lemmatisierte sowie morphosyntaktisch annotierte *Google Books: Spanish* von Mark Davies, dem Ersteller des CdE. Es basiert auf 45 Milliarden Lexemen aus mehreren Millionen Büchern vom 19. bis zum 21. Jahrhundert.⁵⁴⁰ Wie der Name schon vermuten lässt, handelt es sich bei diesem Korpus um Daten aus *Google Books*, wobei es allerdings kein offizielles Produkt von *Google* oder *Google Books* ist, sondern eben von Mark Davies erstellt wurde.⁵⁴¹

Grundsätzlich ist dieses Korpus wie das CdE:Orig (Version 2008) aufgebaut, und es verfügt ebenfalls über die gleiche Suchmaske, doch im Gegensatz zu Letzterem und den anderen Korpora von Davies kann darin weder nach *Key Words in Context* (siehe Kapitel 1.4.3.) gesucht noch können die Treffer zweier Lexeme miteinander verglichen werden. Die Ergebnisse können also lediglich als Liste oder als Grafik angezeigt werden, wobei es sich hier nicht, wie gewöhnlich, um eine Auflistung von Konkordanzen handelt, sondern lediglich um eine Übersicht von Links zu Einträgen von *Google Books*, in denen der jeweilige Suchbegriff vorkommt. Es kann allerdings die Häufigkeit von Lexemen zu verschiedenen Zeiten ermittelt werden.⁵⁴² Dabei werden die Frequenzangaben zum jeweiligen Lexem in Schritten von 10 Jahren unterteilt, das heißt, dass neben der Gesamtfrequenz auch die Häufigkeit des Begriffs von 1810 bis 2000 nach Jahrzehnten aufgelistet wird. Zudem können Kollokationen eines Lexems aufgelistet werden. Wenn

⁵³⁸ Vgl. Davies: *WordAndPhrase*. Op. cit. (20.01.2018).

⁵³⁹ Vgl. Davies, Mark: *New Spanish and Portuguese Corpora*. Online Publikation: <http://corpus.byu.edu/neh2015.asp> (13.07.2019).

⁵⁴⁰ Vgl. Davies: *El corpus del español. Corpora*. Op. cit. (20.01.2018).

⁵⁴¹ Vgl. Davies, Mark: *Google Books: Spanish*. Online Publikation: <https://googlebooks.byu.edu/x.asp> (20.01.2018).

⁵⁴² Vgl. Davies: *El corpus del español. Corpora*. Op. cit. (20.01.2018).

beispielsweise nach Kollokationen des Substantivs *nariz* zusammen mit einem Adjektiv gesucht wird, ermittelt das Korpus folgende Treffer:

The screenshot shows the Google Books: Spanish search interface. On the left, the search string is 'nariz [J*]' and the POS list is 'adj.ALL'. The search results are displayed in a table on the right, listing 18 combinations of 'nariz' with adjectives, their frequency (CHARTS), and total count (TOTAL).

	WORD(S)	CHARTS	TOTAL
1	nariz aguileña	G	12313
2	nariz recta	G	6317
3	nariz larga	G	5264
4	nariz chata	G	4208
5	nariz ancha	G	3325
6	nariz grande	G	2432
7	nariz ganchuda	G	2314
8	nariz prominente	G	2245
9	nariz pequeña	G	1912
10	nariz fina	G	1849
11	nariz corta	G	1819
12	nariz corva	G	1555
13	nariz roja	G	1369
14	nariz aquilina	G	1365
15	nariz rota	G	1196
16	nariz gruesa	G	1103
17	nariz perfilada	G	947
18	nariz pegada	G	863

Abbildung 66: Google Books: Spanish – Suche nach Kollokationen mit dem Substantiv *nariz* und einem Adjektiv.⁵⁴³

Bei einer Suche nach Kollokationen wird das gemeinsame Auftreten des Suchbegriffs zusammen mit einem anderen Lexem ermittelt. Sofern lediglich die Wortart des anderen Lexems angegeben wird, listet das Korpus Kollokationen, bestehend aus der jeweiligen Basis zusammen mit allen Kollokatoren in der angegebenen Wortart, auf. Durch einen Klick auf die jeweilige Kombination gelangt man zu einer Auflistung aller Vorkommen dieser Kollokation in *Google Books*. Dieses Korpus steht prinzipiell kostenlos im Internet zur Verfügung (<https://googlebooks.byu.edu/x.asp>), es kann jedoch, wie alle Korpora von Mark Davies, nur nach einer Registrierung genutzt werden und ist auf 50 Suchanfragen pro Tag beschränkt.

2.6.3. NOW Corpus: Spanish

Im August 2018 wurde ein neues opportunistisches Korpus von Mark Davies veröffentlicht, das *NOW Corpus: Spanish (News on the Web)*. Das *NOW Corpus: Spanish* umfasste im August 2018 mehr als 5,1 Milliarden Lexeme aus 21 verschiedenen spanischsprachigen Ländern, wobei jeden Monat ungefähr 150 Mio. Lexeme in Form von

⁵⁴³ Davies: *Google Books: Spanish*. Op. cit. (20.01.2018).

neuen Artikeln aus Onlinezeitungen und -zeitschriften hinzugefügt werden sollen. Davies unterstreicht daher auf seiner Internetseite, dass man das *NOW Corpus: Spanish* dafür nutzen könne, um Veränderungen der spanischen Sprache bis zum jeweils letzten Monat vor der Suchabfrage zu verfolgen. Damit sei es das einzige Korpus, welches derart aktuelle Nachforschungen zulasse.⁵⁴⁴

Derzeit kann noch nichts darüber ausgesagt werden, ob das Korpus wirklich im angegebenen Maß erweitert werden wird. Im Oktober 2018 enthielt das *NOW Corpus: Spanish* allerdings bereits 5,5 Milliarden Lexeme⁵⁴⁵, im Januar 2019 6,0 Milliarden⁵⁴⁶, im April 2019 6,6 Milliarden⁵⁴⁷, im Juni 2019 6,9 Milliarden⁵⁴⁸ und im Juli 2019 enthält es nun bereits 7,1 Milliarden⁵⁴⁹ Lexeme, was darauf schließen lässt, dass es durchaus kontinuierlich wächst. Vom englischen *NOW Corpus*, welches ebenfalls von Mark Davies erstellt wurde und unter dessen Leitung es kontinuierlich erweitert wird, kann zudem gesagt werden, dass es dieses Versprechen hält. Während es im Januar 2017 noch 3,7 Milliarden Lexeme enthielt⁵⁵⁰, wuchs es bis Oktober 2018 auf 6,5 Milliarden Lexeme.⁵⁵¹ Im Juli 2019 umfasst es nun 8,1 Milliarden Lexeme von Texten aus Onlinezeitungen und -magazinen aus 20 verschiedenen englischsprachigen Ländern von 2010 bis heute.⁵⁵² Es wächst damit tatsächlich, wie angegeben, jeden Monat um 100 bis 120 Mio. Lexeme aus 200.000 bis 250.000 neuen Artikeln beziehungsweise um 1,2 Milliarden Lexeme pro Jahr.⁵⁵³

Die Benutzeroberfläche des *NOW Corpus: Spanish* ist genauso aufgebaut wie diejenige des CdE:Orig (siehe Kapitel 2.1.1.3. und 2.2.1.3.) und des CdE:Nuevo (siehe Kapitel 2.6.1.). Es kann also im Korpus nach Lemmata, Phrasen, Wortarten und Kollokationen gesucht werden. Zudem können die Ergebnisse als Liste oder als Grafik angezeigt werden. Das *NOW Corpus: Spanish* ermöglicht des Weiteren die Entwicklung

⁵⁴⁴ Vgl. Davies: *El corpus del español. Corpora. NOW*. Op. cit. (24.08.2018).

⁵⁴⁵ Vgl. Davies: *Corpus del Español. Corpora*. Op. cit. (06.10.2018).

⁵⁴⁶ Vgl. Davies: *Corpus del Español. Corpora*. Op. cit. (16.01.2019).

⁵⁴⁷ Vgl. Davies: *Corpus del Español. Corpora*. Op. cit. (23.04.2019).

⁵⁴⁸ Vgl. Davies: *Corpus del Español. Corpora*. Op. cit. (19.06.2019).

⁵⁴⁹ Vgl. Davies: *Corpus del Español. Corpora*. Op. cit. (05.07.2019).

⁵⁵⁰ Vgl. Davies, Mark: *NOW Corpus (News on the Web)*. Online Publikation: <http://corpus.byu.edu/now/> (16.01.2017).

⁵⁵¹ Vgl. Davies, Mark: *NOW Corpus (News on the Web)*. Op. cit. (06.10.2018).

⁵⁵² Vgl. Davies, Mark: *NOW Corpus (News on the Web)*. Op. cit. (15.07.2019).

⁵⁵³ Vgl. Davies, Mark: *NOW Corpus (News on the Web)*. Op. cit. (06.10.2018).

virtueller Korpora, um personalisierte Textsammlungen zu erstellen, und lässt geographische wie zeitliche Vergleiche zu.⁵⁵⁴

So kann beispielsweise die Verwendung eines Substantivs mit dem Adjektiv *dulce* in Spanien und Mexiko verglichen werden. Das Korpus liefert hierfür folgende Gegenüberstellung:

SEC 1 (España): 1,145,901,308 PALABRAS							SEC 2 (México): 671,289,375 PALABRAS						
	PALABRA/FRASE	OCURR 1	OCURR 2	P/M 1	P/M 2	PROPORCIÓN		PALABRA/FRASE	OCURR 2	OCURR 1	P/M 2	P/M 1	PROPORCIÓN
1	BOCADO DULCE	37	1	0.0	0.0	21.7	1	VARA DULCE	149	0	0.2	0.0	22.2
2	PUNTO DULCE	72	2	0.1	0.0	21.1	2	CANTANTE DULCE	86	0	0.1	0.0	12.8
3	MOMENTO DULCE	976	30	0.9	0.0	19.1	3	PAN DULCE	490	80	0.7	0.1	10.5
4	PIMENTÓN DULCE	135	7	0.1	0.0	11.3	4	CHILE DULCE	25	7	0.0	0.0	6.1
5	CEBOLLA DULCE	47	3	0.0	0.0	9.2	5	NARANJA DULCE	103	35	0.2	0.0	5.0
6	FRUTA DULCE	183	12	0.2	0.0	8.9	6	CREMA DULCE	23	11	0.0	0.0	3.6
7	MUNDO DULCE	37	3	0.0	0.0	7.2	7	PALO DULCE	23	0	0.0	0.0	3.4
8	JAMÓN DULCE	24	2	0.0	0.0	7.0	8	DIPUTADA DULCE	23	0	0.0	0.0	3.4
9	NOTA DULCE	23	2	0.0	0.0	6.7	9	ACTRIZ DULCE	20	0	0.0	0.0	3.0
10	MUERTE DULCE	63	6	0.1	0.0	6.2	10	FLAUTA DULCE	31	33	0.0	0.0	1.6
11	VERSIÓN DULCE	20	2	0.0	0.0	5.9	11	AGUA DULCE	1273	2132	1.9	1.9	1.0
12	VINO DULCE	258	27	0.2	0.0	5.6	12	SABOR DULCE	178	322	0.3	0.3	0.9

Abbildung 67: *NOW Corpus: Spanish* – Vergleich der Verwendung von *NOUN* + *dulce* in Spanien und Mexiko.⁵⁵⁵

Ebenso kann die Verwendung des Suffixes *-ismo* zur Bildung von Substantiven in den Jahren 2012 bis 2015 mit derjenigen der Jahre 2016 bis 2018 verglichen werden:

SEC 1 (2012-A, 2012-B, 2013-A, 201...): 1,986,550,759 PALABRAS							SEC 2 (2018-B, 2016-A, 2017-A, 201...): 3,330,993,890 PALABRAS						
	PALABRA/FRASE	OCURR 1	OCURR 2	P/M 1	P/M 2	PROPORCIÓN		PALABRA/FRASE	OCURR 2	OCURR 1	P/M 2	P/M 1	PROPORCIÓN
1	SANLUISISMO	103	1	0.1	0.0	172.7	1	ERRATISMO	350	2	0.1	0.0	104.4
2	DERMOGRAFISMO	28	1	0.0	0.0	46.9	2	ERREJONISMO	156	1	0.0	0.0	93.0
3	RASPACUPISMO	25	1	0.0	0.0	41.9	3	ESEVERRISMO	148	1	0.0	0.0	88.3
4	FASISMO	22	1	0.0	0.0	36.9	4	KEIKISMO	113	1	0.0	0.0	67.4
5	PRODISMO	21	1	0.0	0.0	35.2	5	EFRAINISMO	195	3	0.1	0.0	38.8
6	CHAVEZTIALISMO	20	1	0.0	0.0	33.5	6	QUIROGUISMO	64	1	0.0	0.0	38.2
7	SANTOURIBISMO	44	3	0.0	0.0	24.6	7	PROCESISMO	60	1	0.0	0.0	35.8
8	CONSUMERISMO	160	11	0.1	0.0	24.4	8	LLANISMO	658	12	0.2	0.0	32.7
9	MAJUNCHISMO	24	2	0.0	0.0	20.1	9	CARTISMO	1997	47	0.6	0.0	25.3
10	BINITARISMO	23	2	0.0	0.0	19.3	10	TRUMPISMO	169	4	0.1	0.0	25.2
11	NUTRICIONISMO	55	5	0.0	0.0	18.4	11	PPKAUSISMO	42	1	0.0	0.0	25.0
12	ANUMERISMO	22	2	0.0	0.0	18.4	12	PEDRISMO	72	2	0.0	0.0	21.5

Abbildung 68: *NOW Corpus: Spanish* – Vergleich der Substantive mit der Endung *-ismo* von 2012-2015 mit denjenigen von 2016-2018.⁵⁵⁶

Bei dieser Gegenüberstellung sticht insbesondere der Begriff *trumpismo* als zehntes Substantiv in der rechten Auflistung ins Auge. Forscht man in den im *NOW Corpus: Spanish* verfügbaren zeitlichen Abschnitten, so zeigt sich, dass dieser Begriff zum ersten Mal in der zweiten Hälfte des Jahrs 2015 erscheint und das Vorkommen von da an zunimmt. An diesem relativ neuen Begriff ist ersichtlich, dass die Endung *-ismo* zur Bildung von Substantiven im Spanischen produktiv ist.

⁵⁵⁴ Vgl. Davies, Mark: *NOW Corpus: Spanish*. Online Publikation: <https://www.corpusdelespanol.org/now/> (25.08.2018).

⁵⁵⁵ Davies: *NOW Corpus: Spanish*. Op. cit. (25.08.2018).

⁵⁵⁶ Davies: *NOW Corpus: Spanish*. Op. cit. (25.08.2018).

2.7. Persistenz

Dieses Kriterium betrifft die Stabilität der Daten in einem Korpus. Hierbei gilt es, statische und dynamische Korpora zu unterscheiden. Während die meisten Korpora statisch sind, das heißt, eine abgeschlossene Textmenge bilden, die zu einem bestimmten Zeitpunkt gesammelt und dauerhaft gespeichert wurde, werden dynamische Korpora fortlaufend vergrößert oder verändert beziehungsweise aktualisiert.⁵⁵⁷ Diese Korpora werden nie als vollständig angesehen, sondern können zu jeder Zeit durch neue Elemente erweitert werden.⁵⁵⁸ Sie werden daher auch als offen beziehungsweise halboffen bezeichnet, je nachdem, ob der Inhalt vollständig oder nur teilweise erweitert beziehungsweise ausgetauscht wird. Statische Korpora können durchaus auch ergänzt werden, diese Erweiterungen schlagen sich jedoch in einer neuen Version des Korpus nieder, auf die bei Vergleichen oder Nachforschungen stets geachtet werden muss.⁵⁵⁹ Ein Beispiel hierfür ist das in Kapitel 2.1.1.2. beschriebene CORPES, welches sich noch im Aufbau befindet, aber dennoch bereits für eigene Forschungen zur Verfügung steht. Während von Juni 2016 bis September 2018 die Version 0.83 mit 225 Mio. Lexemen genutzt werden konnte⁵⁶⁰, wurde im Juli 2018 die Version 0.9 mit 277 Mio. Lexemen⁵⁶¹ und bereits im Dezember 2018 die Version 0.91 mit 286 Mio. Lexemen veröffentlicht. Mit jeder neuen Version wird das Korpus nicht nur vergrößert, sondern auch verbessert im Hinblick auf Schwierigkeiten früherer Versionen oder sogar um neue Funktionen erweitert.⁵⁶² Je nachdem, welche Version verwendet wird, werden andere Ergebnisse angezeigt, weshalb die Angabe der genutzten Version durchaus von Bedeutung ist.

Sofern sich die Größe der Korpora verändert, spricht man von Monitorkorpora, wobei es hier nebensächlich ist, ob das Korpus kontinuierlich wächst, weil regelmäßig neue Ausgaben einer Tageszeitung ergänzt werden, oder es aus Gründen der Effizienz und des Urheberrechts aktualisiert wird, weil die Texte nicht länger gespeichert werden können.⁵⁶³ Zu beachten ist, dass bei Monitorkorpora in kurzen Zeitabständen neue Texte

⁵⁵⁷ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 105.

⁵⁵⁸ Vgl. Borja (2008): op. cit., S. 246.

⁵⁵⁹ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 105.

⁵⁶⁰ Vgl. Real Academia Española (ed.): *Banco de datos. CORPES*. Op. cit. (24.10.2016).

⁵⁶¹ Vgl. Real Academia Española (ed.): *CORPES*. Op. cit. (27.08.2018).

⁵⁶² Vgl. Real Academia Española (ed.): *CORPES*. Op. cit. (19.06.2019).

⁵⁶³ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 106.

ergänzt und ältere entfernt werden.⁵⁶⁴ Der Nachteil dieser Korpora ist, dass die Ergebnisse der damit durchgeführten Untersuchungen nur bedingt nachprüfbar und wiederholbar sind.

Teubert/Čermáková definieren ein Monitorokorpus wie folgt: „The monitor corpus is a corpus that monitors language change.“⁵⁶⁵ Zudem legen sie dar, dass diese Art von Korpus regelmäßig auf den neuesten Stand gebracht werde und nach oben hin offen sei. Korpuslinguisten seien vor allem interessiert an Veränderungen bezüglich der Häufigkeit von Lexemen und damit auch oft an Veränderungen bezüglich deren Bedeutung und Einteilung in verschiedene Bereiche, am Erscheinen neuer Lexeme und neuer größerer Bedeutungseinheiten sowie an Veränderungen der Kontextprofile (beispielsweise im Hinblick auf die Häufigkeit, mit welcher Ausdrücke im Kontext anderer Lexeme oder Bedeutungseinheiten auftreten). Das Auftreten neuer Lexeme ist die offensichtlichste, aber eine relativ seltene Veränderung, da formal neugebildete Neologismen nur sehr selten vorkommen.⁵⁶⁶ Auch Lemnitzer/Zinsmeister geben an, dass sich Monitorokorpora aufgrund der Tatsache, dass ältere Texte in relativ kurzen Abständen gegen neuere ausgetauscht werden, besonders gut für Untersuchungen eignen, die in kurzen Zeitabständen wiederholt werden, wie in der Lexikographie für die Aufnahme und Beschreibung neuer Lexeme und Wendungen.⁵⁶⁷

Wie bereits erwähnt und wie auch in den anderen Sprachen üblich, sind die meisten spanischen Korpora statisch. Eines der großen Referenzkorpora, das CREA der *Real Academia Española* (siehe Kapitel 2.1.1.1.), ist jedoch ein Monitorokorpus.⁵⁶⁸ Die Texte dieses Korpus stammen aus den letzten 25 Jahren vor dessen Schaffung, also ursprünglich von 1975 bis 1999. Mittlerweile umfasst es Texte von 1975 bis 2004.⁵⁶⁹ Dabei ist die Anzahl der Lexeme allerdings nicht gleichmäßig auf die einzelnen Jahre aufgeteilt, sondern in Zeiträume von jeweils 5 Jahren eingeordnet, wobei den aktuelleren Zeitabschnitten mehr Gewicht verliehen wird.⁵⁷⁰ Die veralteten Texte werden jedoch

⁵⁶⁴ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 189.

⁵⁶⁵ Teubert/ Čermáková (2007): op. cit., S. 71.

⁵⁶⁶ Vgl. Teubert/ Čermáková (2007): op. cit., S. 71-72.

⁵⁶⁷ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 189.

⁵⁶⁸ Vgl. Real Academia Española (ed.): *CREA. Ayuda*. Op. cit. (14.05.2011).

⁵⁶⁹ Vgl. Real Academia Española (ed.): *CREA Escrito y Oral*. Op. cit. (19.05.2011).

⁵⁷⁰ Vgl. Real Academia Española (ed.): *Banco de datos. CREA Escrito*. Online Publikation: <http://www.rae.es/recursos/banco-de-datos/crea-escrito> (14.12.2016).

nicht einfach gelöscht, sondern aus dem CREA herausgenommen und dem CORDE hinzugefügt.⁵⁷¹

Monitorkorpora sollten, wenn möglich, ihre anfängliche Zusammensetzung bewahren, es sollten also Texte verschiedener Genres in dem Maß hinzugefügt werden, dass die Verhältnisse annähernd gleich bleiben.⁵⁷² Dieser Anforderung versucht die *Real Academia Española* im Korpus CREA auf jeden Fall nachzukommen: Um der Aufteilung der Texte des CREA auf die verschiedenen genrespezifischen, chronologischen, geographischen und thematischen Bereiche, welche bereits in Kapitel 2.1.1.1. besprochen wurde, gerecht zu werden, wird versucht, das Gleichgewicht im Design bei der Erweiterung durch neue Texte beizubehalten.⁵⁷³ Allerdings wurde das Korpus seit der Aufnahme der Texte aus den Jahren 2000 bis 2004 nicht mehr aktualisiert, weshalb der Status als Monitorkorpus angezweifelt werden könnte. Zudem ist die chronologische Grenze der neuen, annotierten Version des CREA auf das Jahr 2000 gelegt worden, da das CORPES Texte ab dem Jahr 2001 enthält.⁵⁷⁴ Somit werden die genannten Zweifel vor allem im Hinblick auf die neue Version zunehmend begründeter.

Es sieht jedoch so aus, als ob das *NOW Corpus: Spanish* ein Monitorkorpus der spanischen Sprache werden könnte. Mark Davies deutet auf den Informationsseiten zu seinem Korpus an, dass das *NOW Corpus: Spanish* kontinuierlich mit internetbasierten Texten aus den 21 Ländern, deren Texte es bisher enthält, aktualisiert werden soll. Wie in Kapitel 2.6.3. angegeben, wurde das Korpus erst im August 2018 mit 5,1 Milliarden Lexemen veröffentlicht und umfasst in der derzeitigen Version vom Juli 2019 bereits 7,1 Milliarden Lexeme aus online verfügbaren Zeitungen und Zeitschriften.⁵⁷⁵ Sofern das *NOW Corpus: Spanish* regelmäßig erweitert wird, handelt es sich hierbei tatsächlich um ein Monitorkorpus der spanischen Sprache.

2.8. Verfügbarkeit

Während das Kriterium der Verfügbarkeit von manchen Korpuslinguisten als unerheblich eingestuft wird, ist es für die vorliegende Arbeit nicht zu verachten, denn

⁵⁷¹ Vgl. Real Academia Española (ed.): *CREA. Ayuda*. Op. cit. (12.12.2016).

⁵⁷² Vgl. Borja (2008): op. cit., S. 246.

⁵⁷³ Vgl. Real Academia Española (ed.): *CREA. Ayuda*. Op. cit. (12.12.2016).

⁵⁷⁴ Vgl. Real Academia Española (ed.): *CREA. Versión Anotada*. Op. cit. (03.12.2017).

⁵⁷⁵ Vgl. Davies: *El corpus del español. Corpora. NOW*. Op. cit. (05.07.2019).

kommerziell erhältliche oder nur eingeschränkt verfügbare Korpora sind für die hier vorgestellte Art der Nutzung weniger geeignet.

2.8.1. Im Internet zugängliche Korpora

Einige der bisher vorgestellten Korpora sind im Internet frei zugänglich und kostenlos nutzbar, wie beispielsweise die Korpora der *Real Academia Española* (siehe Kapitel 2.1.1.1., 2.1.1.2., 2.2.1.1. und 2.2.1.2.). Auch die meisten Spezialkorpora, wie das Korpus *Biblia Medieval* (2.2.1.4.), das CORELE (2.2.2.1.), das *Corpus Trilingüe Paralelo Grial* (2.3.1.), das CLUVI (2.3.2.), das CORLEC (2.4.2.3.), das BDS (2.5.2.1.) oder das AnCora (2.5.2.3.) sind frei verfügbar und im Internet nutzbar.

Andere können zwar kostenlos genutzt werden, man muss sich hierfür jedoch registrieren. Dies ist beispielsweise bei den Korpora von Mark Davies der Fall, also bei den Korpora CdE:Orig (siehe Kapitel 2.1.1.3. und 2.2.1.3.), CdE:Nuevo (2.6.1.), *Google Books: Spanish* (2.6.2.) und *NOW Corpus: Spanish* (2.6.3.). Auch das *Spanish in Texas Corpus* (siehe Kapitel 2.1.2.2.) kann erst nach einer Registrierung genutzt werden.

Während viele spanische Korpora zwar online durchsucht werden können, ist der Zugang zu den gesamten Texten bei den meisten nicht möglich. Das Korpus *Hemero* (siehe Kapitel 2.4.1.2.) ist eines der wenigen Korpora, das Links zu den Volltexten bietet. Unter den Korpora von Mark Davies kann zudem das komplette Korpus des CdE:Nuevo seit Dezember 2016 gegen eine Gebühr für den Offline-Gebrauch heruntergeladen werden, wie in Kapitel 2.6.1. zu sehen war.⁵⁷⁶

Außerdem gibt es Korpora, die zwar online verfügbar sind, allerdings nur gegen eine bestimmte Nutzungsgebühr. Beispiele hierfür sind das opportunistische Korpus *esTenTen* und das historische Korpus ADMYTE. Da sich die vorliegende Arbeit allerdings mit Korpora beschäftigt, die für jeden Nutzer frei und kostenlos zugänglich sind, sollen diese Korpora im Folgenden nur kurz erwähnt werden, es wird jedoch im vierten Kapitel nicht auf ihre Nutzbarkeit für den Fremdsprachenunterricht eingegangen.

2.8.1.1. *esTenTen*

Das opportunistische Korpus *esTenTen* ist Teil der Plattform *SketchEngine* und umfasst mehr als neun Milliarden Lexeme.⁵⁷⁷ Das Korpus gesammelter Internettex-

⁵⁷⁶ Vgl. Davies: *Full-text corpus data from corpus.byu.edu – Purchase data*. Op. cit. (13.12.2016).

⁵⁷⁷ Vgl. Lexical Computing (ed.): *Sketch Engine. Corpora*. Online Publikation: <https://the.sketchengine.co.uk/auth/corpora/> (07.05.2018).

2. Kategorisierung der Korpora

besteht aus zwei Subkorpora, dem Webkorpus des europäischen Spanisch (*eseuTenTen*) und dem Webkorpus des amerikanischen Spanisch (*esamTenTen*), wobei der europäische Teil 21% und der amerikanische 79% des Gesamtkorpus einnimmt. Das Korpus *esTenTen* ist lemmatisiert sowie morphosyntaktisch annotiert und verfügt über einige interessante und sehr nützliche Funktionen. Zum einen können, wie bei den meisten anderen Korpora, Konkordanzen und Wortlisten generiert werden, zum anderen bietet es allerdings auch umfassende Darstellungen zu Kollokationen sowie die Möglichkeit, das Korpus als Thesaurus zu nutzen.⁵⁷⁸

Sucht der Nutzer im Korpus *esTenTen* mittels der Funktion *Word sketch* nach Kollokationen, so werden ihm unter anderem frequente Kombinationen aus dem Suchbegriff und Adjektiven, Verben oder anderen Substantiven angezeigt:

pregunta (noun) Alternative PoS: [preguntar verb](#) (freq: 1,694,401)
Spanish Web 2011 (esTenTen11, Eu + Am) freq = [1,393,522](#) (126.85 per million)

modifiers of "pregunta"	17.40	verbs with "pregunta" as object	32.28	verbs with "pregunta" as subject	7.04	"pregunta" and/or ...	5.27
frecuente +	15,506 10.04	responder +	33,977 10.76	surgir +	5,554 9.19	respuesta +	18,353 11.72
preguntas frecuentes		contestar +	21,217 10.26	pregunta que surge		de preguntas y respuestas	
siguiente +	18,740 7.85	formular +	24,097 10.03	acercar +	5,868 8.81	ruegos +	3,647 10.59
la siguiente pregunta :		plantear +	18,006 8.78	preguntas acerca de		. - Ruegos y preguntas .	
retórico +	1,474 7.52	surgir +	11,050 8.65	formular +	1,035 7.98	duda +	3,908 9.83
una pregunta retórica		surge la pregunta		caber +	961 7.85	dudas y preguntas	
incómodo +	1,520 7.48	hacer +	91,066 7.66	pregunta que cabe hacerse es		comentario +	3,895 9.50
preguntas incómodas		obligar +	2,939 7.13	rondar +	492 6.95	preguntas y comentarios	
obvio +	1,635 7.45	pregunta obligada		pregunta que ronda		ruego +	1,594 9.40
pregunta obvia		caber +	2,274 6.95	molestar +	406 6.88	de ruegos y preguntas	
tonto +	1,393 7.36	cabe la pregunta		pregunta no molesta .		inquietud +	1,748 8.84
una pregunta tonta		relacionar +	5,374 6.77	plantear +	905 6.70	preguntas e inquietudes	
difícil +	3,334 7.23	preguntas relacionadas con		guiar +	353 6.59	sugerencia +	942 7.93
pregunta difícil		lanzar +	3,151 6.57	apuntar +	757 6.52	preguntas y sugerencias	
sencillo +	2,531 7.12	repetir +	2,231 6.47	pregunta apunta		observación +	849 7.78
sencilla pregunta		escuchar +	2,640 6.26	quedar +	2,448 6.50	de preguntas u observaciones	
concreto +	3,198 6.95	venir +	2,319 6.23	pregunta que queda		respuestas +	522 7.76
preguntas concretas		viene la pregunta		flotar +	296 6.48	. Preguntas y Respuestas	
interesante +	2,816 6.90			pregunta que flota en el		pregunta +	900 7.65

Abbildung 69: *esTenTen* – Kollokationen zum Substantiv *pregunta*.⁵⁷⁹

In dieser Funktion ist die Frequenz der Kollokation sowie die Art der Kombination, also beispielsweise die Position des Adjektivs vor oder nach dem Substantiv, unmittelbar ersichtlich. Zudem gelangt der Nutzer mittels eines Klicks auf einen Kollokator zu einer Auflistung der Konkordanzen.

⁵⁷⁸ Vgl. Lexical Computing (ed.): *Sketch Engine. esTenTen*. Online Publikation: <https://www.sketchengine.eu/esTenTen-spanish-corpus/> (07.05.2018).

⁵⁷⁹ Lexical Computing (ed.): *Sketch Engine. esTenTen. Word sketch: pregunta*. Online Publikation: https://the.sketchengine.co.uk/corpus/wsketch?corpname=preloaded%2Festenten11_fl5&reload=&lemma=pregunta&lpos=-n&usesubcorp=&minfreq=auto&minscore=0.0&maxitems=25&sort_ws_columns=s&show_lemma_coverage=0&show_lcm=0&show_lcm=1&clustercols=0&minsim=0.15&structured=0&structured=1&min_unary_score=0.0&min_mwlink_freq=100&nr_ws_cols=5&bim_lang= (07.05.2018).

2.8.1.2. ADMYTE – Archivo Digital de Manuscritos y Textos Españoles

Das *Archivo Digital de Manuscritos y Textos Españoles* (kurz: ADMYTE), ist die Arbeit einer Gruppe von Sprachhistorikern des *Historic Seminary of Medieval Studies* der *University of Wisconsin-Madison*. In der aktuellen Version umfasst das Projekt vollständige Transkriptionen von 290 Werken des Spanischen oder eines seiner Dialekte im Mittelalter und damit mehr als 54.000 Seiten.⁵⁸³

Eine erste Version (ADMYTE 0) wurde im Jahr 1991 veröffentlicht, enthielt jedoch lediglich 64 mittelalterliche Texte. Die zweite Version (ADMYTE I) umfasste Transkriptionen und Bilder von 61 Werken aus der Zeit der Eroberung Amerikas. Die neue Version (ADMYTE II) enthält nun 165 Texte, so auch grundlegende Werke der Renaissance wie den *Lazarillo de Tormes* oder des Barocks wie das Manuskript *Chacón* mit den Werken der Dichtkunst von Don Luis de Góngora.⁵⁸⁴

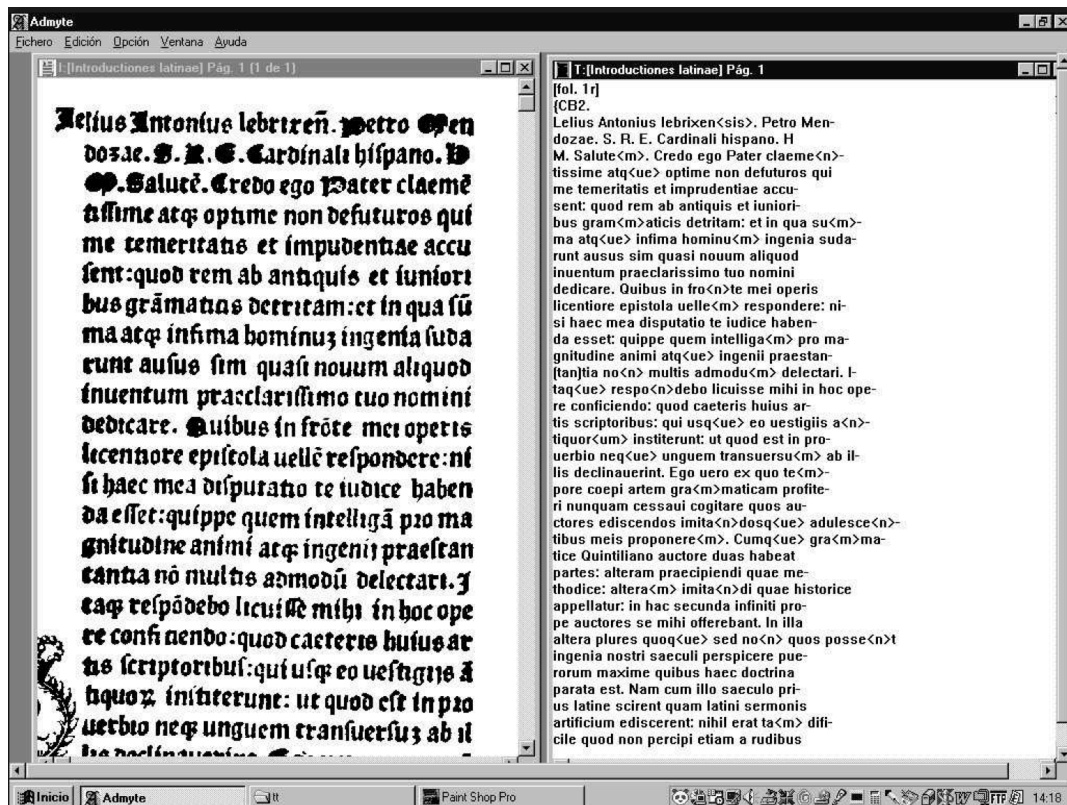


Abbildung 71: ADMYTE – Auszug aus dem Korpus ADMYTE.⁵⁸⁵

⁵⁸³ Vgl. Marcos Marín, Francisco/ Faulhaber, Charles B.: *ADMYTE. Presentación*. Online Publikation: <http://www.admyte.com/admyteonline/presentacion.htm> (27.10.2016).

⁵⁸⁴ Vgl. Marcos Marín, Francisco/ Faulhaber, Charles B.: *ADMYTE. Historia*. Online Publikation: <http://www.admyte.com/admyteonline/historia.htm> (27.10.2016).

⁵⁸⁵ Marcos Marín, Francisco (1998): *La Colección Foulché-Delbosc de la Biblioteca Nacional de la República Argentina. Catalogación y Edición Electrónica*. Online Publikation: <http://www.illf.uam.es/~fmarcos/informes/BNArgentina/fd01.htm> (06.06.2017).

Das aktuelle Korpus ADMYTE enthält viele mittelalterliche Meisterwerke wie *Cantar de Mio Cid*, *Libro de Alexandre*, *Libro de Buen Amor* oder die Tragikomödie *Calisto y Melibea* von Fernando de Rojas, aber auch Enzyklopädien, Wörterbücher, Grammatiken, Romane, Liederbücher oder wissenschaftliche Abhandlungen sind darunter. Es wird vom Ersteller, Francisco Marcos Marín, als das derzeit größte Textkorpus, auf das Sprachhistoriker mit dem Schwerpunkt Mittelalter zurückgreifen können, gelobt und bedarf laut der Entwickler keinerlei Einführung, sondern ist intuitiv bedienbar.⁵⁸⁶

Nachdem das Korpus ADMYTE anfangs lediglich auf CD-ROM erhältlich war, ist die neue Version nun im Internet verfügbar (www.admyte.com). Allerdings muss für die Nutzung eine relativ hohe Gebühr entrichtet werden. Einzelpersonen (Forscher/Studenten) zahlen für einen Zugang pro Jahr 100 €, Institutionen für 15 Zugänge 1.000 € und Korporationen für eine unbegrenzte Anzahl an Zugängen 12.000 €. Auf diese Beträge fallen zusätzlich noch Steuern in Spanien an.⁵⁸⁷

2.8.2. Als CD-ROM verfügbare Korpora

Neben diesen kostenpflichtigen Korpora, die im Internet zugänglich sind, gibt es manche Korpora lediglich auf CD-ROM. Dies ist der Fall bei den bereits vorgestellten Korpora des gesprochenen Spanisch *Macrocorpus* (siehe Kapitel 2.4.2.1.) und C-ORAL-ROM (siehe Kapitel 2.4.2.2.). Der Vorteil dieser beiden Korpora ist, dass auf der CD-ROM alle Texte verfügbar sind, man also die kompletten Texte nutzen kann, sofern man den Datenträger vorliegen hat. Die beiden Korpora bleiben jedoch auf dem Stand der Veröffentlichung und werden nicht aktualisiert wie dies bei online zugänglichen Korpora in der Regel der Fall ist.

Einige andere Korpora sind ebenso wenig im Internet zugänglich, da sie nur auf CD-ROM verfügbar sind, das heißt, man müsste sie käuflich erwerben oder in einer Bibliothek ausleihen, um darauf zurückgreifen zu können. Daher sollen diese Korpora im Folgenden nur kurz erwähnt werden, es wird jedoch, wie im Fall der beitragspflichtigen Korpora in Kapitel 2.8.1., im Verlauf der Arbeit nicht näher auf mögliche

⁵⁸⁶ Vgl. Marcos Marín, Francisco/ Faulhaber, Charles B.: *ADMYTE. Contenido*. Online Publikation: <http://www.admyte.com/admyteonline/contenido.htm> (27.10.2016).

⁵⁸⁷ Vgl. E-Mail-Kontakt mit Juan Manuel García von *Micronet* am 14.03.2017.

Herangehensweisen oder ihre Nutzbarkeit eingegangen, da sich die vorliegende Arbeit hauptsächlich mit Korpora beschäftigen soll, die frei und kostenlos verfügbar sind.

2.8.2.1. PRESEEA – *Proyecto para el Estudio Sociolingüístico del Español de España y de América*

Das *Proyecto para el Estudio Sociolingüístico del Español de España y de América* (PRESEEA) ist ein im Jahr 1992 initiiertes und von Francisco Moreno Fernández geleitetes Projekt zur Erstellung eines Korpus der gesprochenen spanischen Sprache, das in seiner geographischen und sozialen Streuung für die spanischsprachige Welt repräsentativ sein soll. Es wird von 40 soziolinguistischen Forschergruppen vorangebracht, deren Ziel es ist, eine für Bildung und Technologie nutzbare Datenbank der soziolinguistischen Vielfalt der spanischsprachigen Gemeinschaft zu erstellen.⁵⁸⁸

Die Informanten werden nach drei soziolinguistischen Gesichtspunkten eingeteilt (Geschlecht, Alter, Bildungsgrad) und nehmen an Interviews und Gesprächen über Alltagsthemen teil.⁵⁸⁹ Die Daten werden auf CD-ROM⁵⁹⁰ zur Verfügung gestellt und einige Audiodateien aus unterschiedlichen spanischsprachigen Ländern sowie deren Transkriptionen können zudem im Internet heruntergeladen werden. Dort werden Interessierte auch regelmäßig über Fortschritte informiert (<http://preseea.linguas.net/Inicio.aspx>).⁵⁹¹

Die im Internet zur Verfügung gestellten Dateien enthalten ganze Interviews von beispielsweise zehn Minuten Länge sowie deren Transkription und können von den Nutzern frei verwendet werden.⁵⁹² Die Interviews können nach Städten, Geschlecht des Sprechers, Altersgruppe und Bildungsniveau sortiert, es kann jedoch nicht speziell nach Themen oder einzelnen Lexemen gesucht werden. Da es sich um ein Projekt zur Untersuchung des gesprochenen Spanisch handelt, enthält die Transkription viele

⁵⁸⁸ Vgl. Moreno Fernández, Francisco: *Proyecto para el Estudio Sociolingüístico del Español de España y de América. Presentación*. Op. cit. (27.10.2016).

⁵⁸⁹ Vgl. Becker (2013): op. cit., S. 203-204.

⁵⁹⁰ Vgl. Moreno Fernández, Francisco (2002): *La lengua hablada en Alcalá de Henares. Corpus PRESEEA-Alcalá*. Alcalá: Univ. de Alcalá. CD-ROM.

⁵⁹¹ Vgl. Moreno Fernández, Francisco: *Proyecto para el Estudio Sociolingüístico del Español de España y de América. Consulta Corpus PRESEEA*. Online Publikation: <http://preseea.linguas.net/Corpus.aspx> (27.10.2016).

⁵⁹² Vgl. PRESEEA (ed.2014-2017): *Corpus del Proyecto para el estudio sociolingüístico del español de España y de América*. Alcalá de Henares: Universidad de Alcalá. Online Publikation: <http://preseea.linguas.net> (09.01.2017).

Informationen zur Sprechweise der Informanten, wie beispielsweise Zögern, längere Pausen oder Lachen.

```
<Habla nte id="hab3" nombre="María Sancho" codigo_hab="A1" sexo="mujer" grupo_edad="1"
edad="20" nivel_edu="alto" estudios="segundo curso filología" profesion="estudiante"
origen="Madrid"

papel="audiencia"/>

<Relaciones rel_ent_inf="desconocidos" rel_inf_aud1="desconocidos"

rel_ent_aud1="conocidos" rel_inf_aud2="no" rel_ent_aud2="no"/>

</Habla ntes> </Trans>
```

```
E: pues vamos a empezar primero le voy a hacer unas preguntas<alargamiento/> /
generales y luego ya empezamos con la encuesta ver· que son cosas /

I: sí

E: generales que no<alargamiento/>

I: pero ¿esto es de la universidad de Alcalá de Henares?

E: de de la Universidad de Alcalá.

I: ¿del del departamento de sociología o así?

E: eeh no estamos haciendo un estudio general sociológico de historia pero no es
<alargamiento/> el departamento de sociología
```

Abbildung 72: PRESEEA – Ausschnitt aus einem Interview mit einer Frau der oberen Bildungsschicht aus Madrid.⁵⁹³

In obigem Ausschnitt ist zu erkennen, dass zunächst in jeder Datei wichtige Informationen zum Interview festgehalten werden. Dabei ist zu erwähnen, dass der Beginn dieses Ausschnitts nicht dem Anfang der Datei entspricht. Es finden sich weitere Informationen zum Hintergrund des Interviews am Beginn der Datei. Die zweite Hälfte des Ausschnitts stellt den Anfang des Interviews dar, wobei die eben beschriebenen Informationen zur Sprechweise der Informanten beispielsweise anhand des Einschubs *<alargamiento/>* zu erkennen sind.

2.8.2.2. LEXESP – *Léxico informatizado del español*

Ein weiteres Korpus, welches lediglich auf CD-ROM veröffentlicht wurde, ist das Korpus LEXESP der *Universitat de Barcelona*. Dieses Korpus wurde entwickelt, weil Wissenschaftler aus verschiedensten Fachbereichen, unter anderem der Psychologie, der Sprachwissenschaft, der Medizin und der Erziehungswissenschaften, welche sich mit bestimmten sprachlichen Aspekten beschäftigen, eine schnelle und effiziente

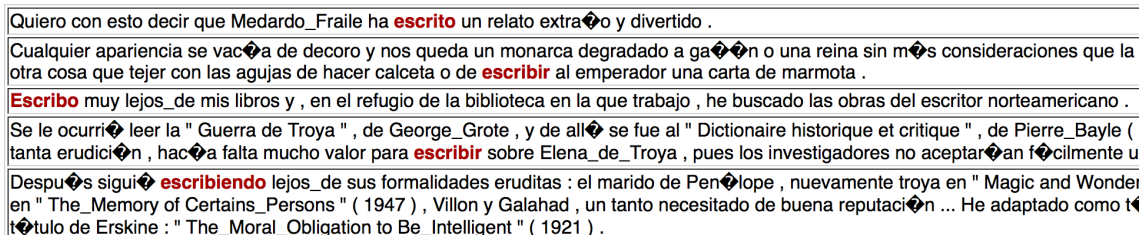
⁵⁹³ PRESEEA (ed.2014-2017): *Corpus del Proyecto para el estudio sociolingüístico del español de España y de América. Corpus*. Alcalá de Henares: Universidad de Alcalá. Online Publikation: <http://preseea.linguas.net/Corpus.aspx> (14.03.2017).

2. Kategorisierung der Korpora

Informationsquelle für spezifische Eigenschaften von Lexemen benötigten. Die Autoren begründen die Erstellung dieses Korpus damit, dass bis zu dessen Erscheinen lediglich Wörterbücher sowie verschiedene Wortlisten verfügbar waren, wodurch zum einen die Suchen manuell, langsam und fehlerbehaftet waren und zum anderen die Informationen in verschiedene Teile zerlegt und verstreut waren.⁵⁹⁴

Das LEXESP besteht aus zwei Teilen: Eine Komponente nutzt das lemmatisierte Korpus als Basis für Suchen nach Lexemen, Kookkurrenzen, Lemmata und Wortarten in einem gewissen Kontext. Die zweite Komponente verwendet Wortlisten für Studien, welche die Suche nach besonderen Eigenschaften von Lexemen ermöglichen. Die CD-ROM enthält zudem drei bereits durchgeführte Studien, welche vom Nutzer angepasst werden können.⁵⁹⁵ Das LEXESP umfasst 5,5 Mio. Lexeme des aktuellen Spanisch und enthält unter anderem verschiedene literarische Genres, Nachrichten, Zeitungsartikel und wissenschaftliche Texte.⁵⁹⁶

Dieses Korpus zeichnet sich durch die Besonderheit aus, dass zu Demonstrationszwecken eine Internetseite der Universität Oviedo gestaltet wurde, mittels derer exemplarisch im Korpus gesucht werden kann (<http://www.cs.upc.edu/~nlp/tools/corpus-es.php>).⁵⁹⁷ Die Seite ist zwar nur spärlich gestaltet, doch bei der Suche kann sogar die Wortart angegeben werden.



Quiero con esto decir que Medardo_Fraile ha **escrito** un relato extraño y divertido .

Cualquier apariencia se vacía de decoro y nos queda un monarca degradado a galeón o una reina sin más consideraciones que la otra cosa que tejer con las agujas de hacer calceta o de **escribir** al emperador una carta de marmota .

Escribo muy lejos de mis libros y , en el refugio de la biblioteca en la que trabajo , he buscado las obras del escritor norteamericano .

Se le ocurrió leer la " Guerra de Troya " , de George Grote , y de allí se fue al " Dictionaire historique et critique " , de Pierre Bayle (tanta erudición , hacía falta mucho valor para **escribir** sobre Elena de Troya , pues los investigadores no aceptaban fácilmente u

Después siguió **escribiendo** lejos de sus formalidades eruditas : el marido de Penélope , nuevamente troya en " Magic and Wonder en " The Memory of Certain Persons " (1947) , Villon y Galahad , un tanto necesitado de buena reputación ... He adaptado como título de Erskine : " The Moral Obligation to Be Intelligent " (1921) .

Abbildung 73: LEXESP – Suche nach dem Lemma *escribir*.⁵⁹⁸

Wie man an obigem Ausschnitt erkennen kann, ist es möglich im LEXESP nach Lemmata zu suchen. Gibt man nun das Lemma *escribir* und die Wortart V für *verbo* ein, so ermittelt die Suchmaschine alle Vorkommen der Formen von *escribir* im Korpus.

⁵⁹⁴ Vgl. Sebastián Gallés, Núria/ Cuetos Vega, Fernando/ Carreiras Valiña, Manuel Francisco/ Martí Antonin, Maria Antònia (2000): *LEXESP. Léxico informatizado del español*. Barcelona: Edicions de la Universitat de Barcelona, S. 7.

⁵⁹⁵ Vgl. Universitat de Barcelona (ed.2016): *LEXESP. Léxico informatizado del español*. Online Publikation: <http://www.publicacions.ub.edu/ficha.aspx?cod=05096e> (22.12.2016).

⁵⁹⁶ Vgl. Uniovi (ed.): *Spanish Corpus Search*. Online Publikation: <http://www.cs.upc.edu/~nlp/tools/corpus-es.php> (27.01.2017).

⁵⁹⁷ Vgl. Uniovi (ed.): *Spanish Corpus Search*. Op. cit. (27.01.2017).

⁵⁹⁸ Uniovi (ed.): *Spanish Corpus Search*. Op. cit. (14.03.2017).

Allerdings hat die Seite Schwierigkeiten bei der Darstellung der spanischen Sonderzeichen, weshalb diese bei der Verwendung der Beispielsätze noch ergänzt werden müssen.

2.9. Das Internet – Ein Korpus?

In Zusammenhang mit den einzelnen Korpora, den darin enthaltenen Texten und den sich dort bietenden Suchoptionen stellt sich häufig die Frage, ob das Internet nicht auch ein Korpus sei. Da diese Frage nicht ohne Weiteres in einer kurzen Antwort begründet werden kann, soll sich der letzte Teil des vorliegenden Kapitels mit der Lösung derselben beschäftigen.

Die Zuschreibung des Internets zu den Korpora ist verständlicherweise abhängig von der Definition Letzterer. Sofern man ein Korpus lediglich als Sammlung von Texten zum Zweck sprachwissenschaftlicher Untersuchung versteht, ist das Internet laut Scherer als solches zu bezeichnen. Wird es allerdings als Textsammlung gesehen, welche nach sprachwissenschaftlichen Kriterien zusammengestellt wurde und bestimmten Ansprüchen wie der Repräsentativität oder der Beständigkeit genügen muss, so ist es aufgrund seiner nicht zu erfassenden Größe und seiner kontinuierlichen Veränderung wohl nicht als solches zu definieren.⁵⁹⁹ Für die vorliegende Arbeit wurde im ersten Kapitel das Prinzip der Repräsentativität als verbindliches Kriterium für die Korpusdefinition angegeben. Daher stellt das Internet laut dieser Definition kein Korpus dar. Dennoch gibt es gute Gründe, um das *World Wide Web* durchaus als Korpus zu kategorisieren beziehungsweise es zumindest ähnlich einem Korpus zu nutzen. Hinsichtlich der Einordnung des Internets als Korpus existieren die unterschiedlichsten Ansichten, wie anhand der im Folgenden aufgelisteten Begründungen einiger Korpuslinguisten zu sehen sein wird.

Leech sieht das Internet auch nicht als Korpus, da er ebenso die Repräsentativität desselben infrage stellt. Aufgrund der Tatsache, dass Suchmaschinen wie *Google* Algorithmen verwenden, deren Funktionsweise für den gewöhnlichen Nutzer unerklärlich sind, sie keine Suchen im gesamten *World Wide Web* zulassen und ihre Ergebnisse instabil und inkonsistent sind, erhält man beim Durchforsten des Internets lediglich eine Art umfassende Stichprobe. Wie repräsentativ diese Stichprobe ist, bleibt

⁵⁹⁹ Vgl. Scherer (2006): op. cit., S. 74.

ein Rätsel. Zudem liefern Suchmaschinen keine repräsentativen Ergebnisse zum Sprachgebrauch, da manche Genres wie akademische Texte überrepräsentiert werden, während andere Bereiche wie die gesprochene Sprache oder Manuskripte unterrepräsentiert sind. Das Netz kann uns zwar aufschlussreiche Daten zu bestimmten Untersuchungen bereitstellen, diese sind jedoch nur provisorisch und müssen genauer geprüft werden, da bei den meisten Internetseiten nicht einmal bekannt ist, ob sie von Muttersprachlern verfasst wurden und aus welchem Land oder welcher Region sie stammen.⁶⁰⁰

Auf der anderen Seite kommen Kilgariff/Grefenstette nach genauer Analyse der Repräsentativität des Internets zu dem Entschluss, dass es nur für sich selbst repräsentativ ist, so wie jedoch auch alle anderen Korpora. Sie sind der Meinung, in Bezug auf unser Textverständnis und die Frage nach der Repräsentativität oder Ausgeglichenheit eines Korpus stecken wir noch in den Kinderschuhen und jede genauere Untersuchung führe zu weiteren Fragen. Sofern wir nicht wissen, welche Texte in welchem Verhältnis für welche Sprachvarietät repräsentativ sind, seien das Internet und alle Korpora zumindest in diesem Punkt gleichzusetzen.⁶⁰¹

Selbst wenn man Eigenschaften wie die Repräsentativität beiseitelässt, so sollte ein Korpus aber zumindest den realen Sprachgebrauch einer Sprachgemeinschaft darstellen. Dies kann durchaus mithilfe des Internets geschehen, sofern sich die Untersuchungen auf Neologismen, neue Verwendungsweisen von Lexemen und Kollokationen beziehen, da das Netz hier aufgrund seiner immensen Datenmenge jedes Korpus schlägt. Die fehlenden linguistischen Annotationen wie beispielweise das *POS-tagging* oder die Lemmatisierung haben jedoch zur Folge, dass grammatische und semantische Untersuchungen begrenzt sind. Erst wenn es mithilfe intelligenter Suchmaschinen möglich ist, diese Hürde zu meistern, rücken wir laut Leech der Annahme näher, das Internet als Korpus ansehen zu können.⁶⁰² Diese fehlenden linguistischen Annotationen sind ein eindeutiger Minuspunkt des Internets im Vergleich zu Korpora, da Flexion und Wortstellung nicht berücksichtigt werden. Will man die generelle Möglichkeit der Nutzung eines Lexems in einer bestimmten Kombination ermitteln, so muss man einzeln

⁶⁰⁰ Vgl. Leech (2007): op. cit., S. 144-145.

⁶⁰¹ Vgl. Kilgariff, Adam/ Grefenstette, Gregory (2003): „Web as Corpus“. In: *Computational Linguistics* 3/29. Online Publikation: <https://www.kilgariff.co.uk/Publications/2003-KilgGrefenstette-WACIntro.pdf> (11.11.2016), S. 8-11.

⁶⁰² Vgl. Leech (2007): op. cit., S. 145.

nach allen Formen des Lexems suchen, um genau zu wissen, welche Verwendungen möglich sind. Eine Suche mit dem Lemma, bei der man alle existenten Formen als Ergebnis erhält, ist nicht möglich.

Auch Gerstenberg sieht das Internet im korpuslinguistischen Sinn nicht als Korpus. Zur Untermauerung führt sie Kilgariffs Aufsatz „Googleology is bad science“ aus dem Jahr 2006 an, welcher auf den Qualitätsverlust durch unreflektierte Suchen im Internet aufmerksam macht.⁶⁰³ In der bereits erwähnten Veröffentlichung von Kilgariff/Grefenstette aus dem Jahr 2003 ordneten die Autoren das *World Wide Web* aufgrund einer weiten Korpusdefinition den Korpora zu, da man die Frage ‚Was ist ein Korpus?‘ nicht mit der Frage ‚Was ist ein gutes Korpus?‘ verwechseln dürfe, wie dies ihrer Meinung nach in der Vergangenheit häufig der Fall gewesen sei. Daher definierten sie ein Korpus als eine Sammlung von Texten, wenn diese als Objekt sprachlicher oder literarischer Untersuchung gilt, und bejahten die Frage nach dem Internet als Korpus somit.⁶⁰⁴ Sie sind der Meinung, das *World Wide Web* könne durchaus als multilinguales Korpus gesehen werden.⁶⁰⁵ In seinem späteren Aufsatz aus dem Jahr 2006 betrachtet Kilgariff die Verwendung des Internets jedoch zunehmend kritisch und bemängelt die Tatsache, dass Suchmaschinen teilweise die Zahl der Abfragen begrenzten und zudem lediglich über eine beschränkte Suchsyntax verfügten. Darüber hinaus ergäben identische Suchanfragen an aufeinanderfolgenden Tagen bisweilen abweichende Ergebnisse, da die Anfragen an verschiedene Computer zu unterschiedlichen Zeitpunkten im Aktualisierungsprozess geschickt würden und daher andere Resultate lieferten. Er sei zwar immer noch der Meinung, dass das Internet durchaus zu linguistischen Zwecken genutzt werden sollte, da es umfassend, frei und ständig verfügbar sei sowie in weiten Teilen sprachwissenschaftlichen Kriterien genüge, betont aber, dass Nutzer, wenn sie die bereits erwähnten Nachteile in Kauf nähmen und die Algorithmen und Besonderheiten der Suchmaschinen genauer erforschten, keine Experten der Sprachwissenschaft, sondern einer anderen Wissenschaft, der von ihm so betitelten ‚Googleology‘ würden.⁶⁰⁶

⁶⁰³ Vgl. Gerstenberg (2013): op. cit., S. 95.

⁶⁰⁴ Vgl. Kilgariff/ Grefenstette (2003): op. cit., S. 2.

⁶⁰⁵ Vgl. Kilgariff/ Grefenstette (2003): op. cit., S. 5.

⁶⁰⁶ Vgl. Kilgariff, Adam (2006): „Googleology is Bad Science“. *Computational Linguistics* 1/33, S. 147-148. Online Publikation: https://www.sketchengine.co.uk/wp-content/uploads/Googleology_bad_science_2007.pdf (11.11.2016).

2. Kategorisierung der Korpora

Bereits in ihrem früheren Aufsatz bemängeln Kilgarriff/Grefenstette die Funktionsweise von Suchmaschinen wie *Google* in Bezug auf linguistische Untersuchungen. Zunächst werden nicht genügend Treffer angezeigt (nur maximal 5000), es wird nicht genügend Kontext angegeben (bei *Google* nur ungefähr 10 Lexeme), die Auswahlkriterien sind aus linguistischer Sicht verzerrt, die Suchen können nicht nach linguistischen Kriterien spezifiziert werden und die Statistiken sind unzuverlässig, weil die Anzahl der Seiten, auf denen der Ausdruck steht, gezählt wird und nicht die gesamte Häufigkeit.⁶⁰⁷

Zur Qualität der Daten ist darüber hinaus zu sagen, dass die Texte eine unbekannte Anzahl an falschen Lexemen und Ausdrücken enthalten, welche dennoch beibehalten und gegebenenfalls als Treffer angezeigt werden.⁶⁰⁸ So finden sich neben 300 Mio. Treffern für das englische Verb *receive* auch knapp neun Mio. Treffer für die inkorrekte Form **recieve*. McEnery/Hardie sehen hierin allerdings die Möglichkeit häufige Rechtschreibfehler über das Internet zu ermitteln und diese als Basis für eine Studie zur Unterstützung einer Rechtschreibreform zu nutzen. Sollte man jedoch nicht an derartigen Studien interessiert sein, müsse man sich überlegen, wie man diesen Fehleranteil, der sich durch das gesamte Netz zieht, umgeht. Nichtsdestotrotz stellt das Internet für sie eine wichtige Quelle für eine Vielzahl von Daten dar, welche ausgewählt und für die Erstellung von Korpora zu verschiedensten Zwecken verwendet werden können.⁶⁰⁹

McEnery/Hardie sehen das Konzept des Internets als Korpus ähnlich wie die Idee des Monitorkorpus (siehe Kapitel 2.7.). Es stellt eine große, ständig wachsende Datensammlung dar, welche für sprachwissenschaftliche Untersuchungen verwendet werden kann. Neben der Suche mit gewöhnlichen Suchmaschinen wie *Google* haben Forscher spezielle Benutzeroberflächen erarbeitet, um diese Art der Verwendung des Internets zu unterstützen.⁶¹⁰ Dabei bieten Konkordanzprogramme wie die Internetkonkordanz *WebConc* von Matthias Hüning eine gute Möglichkeit, um die Internettexpte nach bestimmten Suchausdrücken zu durchforsten.⁶¹¹

Dennoch darf man die Einwände von McEnery/Hardie bezüglich der Richtigkeit der Daten sowohl in Bezug auf die Wahrhaftigkeit der Informationen als auch hinsichtlich

⁶⁰⁷ Vgl. Kilgarriff/ Grefenstette (2003): op. cit., S. 12-13.

⁶⁰⁸ Vgl. Scherer (2006): op. cit., S. 74-75.

⁶⁰⁹ Vgl. McEnery/ Hardie (2012): op. cit., S. 7-8.

⁶¹⁰ Vgl. McEnery/ Hardie (2012): op. cit., S. 7-8.

⁶¹¹ Vgl. Scherer (2006): op. cit., S. 76.

der Rechtschreibung nicht verachten. Jeder Nutzer kann Informationen in Form von Texten ins Internet stellen, ganz gleich, ob es sich dabei um korrekte oder inkorrekte Inhalte handelt und ob der Verfasser den Text in seiner Muttersprache oder einer Fremdsprache verfasst hat. Somit warten zahlreiche inhaltlich und formal falsche Informationen auf Internetseiten, welche vor allem in der Fremdsprache nicht immer sofort von korrekten und fundierten Informationen unterschieden werden können.

Aus Sicht der Übersetzungswissenschaft sieht Anabel Borja das Internet ebenso als eine Art Korpus. Sie sieht Korpora als große Textsammlungen in elektronischer Form und beschreibt das Internet in Anbetracht dieser weiten Definition als ein dynamisches, offenes, frei zugängliches und sich kontinuierlich entwickelndes Korpus. Es ist das größte existierende Korpus und umfasst den breitesten Anwendungsbereich.⁶¹² Teubert/Čermáková bestätigen, dass auch Korpuslinguisten das Internet durchaus als virtuelles Korpus nutzen. Es wird vor allem zur Überprüfung der Existenz von Lexemen oder Phrasen und deren jeweiligen Kontexten verwendet, wenn diese nicht in den klassischen Korpora zu finden sind.⁶¹³

Auch McEnergy/Hardie sehen einen großen Vorteil des Internets in der Tatsache, dass Lexeme oder Phrasen, welche in Korpora nur relativ selten vorkommen, eine viel höhere Trefferzahl im Internet hervorbringen, auf die der Wissenschaftler seine Beobachtungen stützen kann. Zugegebenermaßen müssen die Treffer aus dem Internet erst überprüft und sortiert werden, sie bieten aber dennoch ein viel genaueres und nuancierteres Bild des zu untersuchenden Ausdrucks. Nichtsdestotrotz liegt in diesem Vorteil auch ein Nachteil: Bei Suchanfragen nach häufig vorkommenden Ausdrücken muss die enorme Menge an Beispielen im Internet erst einmal sortiert und viele Treffer müssen beiseitegelassen werden. Diese Auswahl sollte einer heuristischen Methode folgen, welche konsistent bei allen Analysen angewandt werden muss.⁶¹⁴

Denn auch wenn die Suche auf die frei zugänglichen Webseiten beschränkt wird, ist das Internet, wenn nicht unendlich, so doch sicherlich unerschöpflich. Kein Browser kann behaupten alle existierenden Webseiten abzudecken. Eine solche Auswahl ist gewöhnlich so umfassend, dass in dem Moment, in dem alle Treffer für einen Suchbegriff

⁶¹² Vgl. Borja (2008): op. cit., S. 243-245.

⁶¹³ Vgl. Teubert/ Čermáková (2007): op. cit., S. 76-77.

⁶¹⁴ Vgl. McEnergy/ Hardie (2012): op. cit., S. 8.

extrahiert sind, einige gesuchte Texte bereits vom Server genommen und andere, neuere hinzugefügt worden sind.⁶¹⁵ Wenn die für eine Studie verwendeten Daten aus dem Internet nicht heruntergeladen und angemessen gespeichert wurden, ist es schwerlich möglich, diese Studie nach einigen Jahren nachzuprüfen, da sich das Internet ständig verändert. Aufgrund der Tatsache, dass die Nachvollziehbarkeit und Überprüfbarkeit solch experimenteller Vorhaben eine signifikante Rolle spielt, sind die Vergänglichkeit und der Wandel ein offensichtlicher und schwerwiegender Nachteil, der für McEnery/Hardie gegen den Ansatz des Internets als Korpus spricht.⁶¹⁶ Teubert/Čermáková meinen hierzu, das Internet sei ein virtuelles Korpus und, wie für die Sprache jeglicher Sprachgemeinschaft zutreffend, könne man nicht davon ausgehen, diese in ihrer Gesamtheit zu erforschen.⁶¹⁷

Bezüglich der Größe des *World Wide Web* können nur Schätzungen vorgenommen werden. Im Jahr 2003 wurde die Größe des deutschsprachigen Teils auf sieben Milliarden Lexeme geschätzt, die des englischsprachigen Teils auf 75 Milliarden. Problematisch sieht Scherer hierbei, dass die Ergebnisse von Korpusanalysen meist in Relation zur Korpusgröße angegeben werden, um damit einen Bezugsrahmen zu schaffen. Diese Einordnung und damit die Aussagekraft der Suchergebnisse sind beim Internet nicht gegeben. Demnach erlaubt das Internet zwar qualitative, aber keine quantitativen Analysen.⁶¹⁸

Da das Internet größer ist als jede existierende Bibliothek, kann davon ausgegangen werden, dass ein Lexem, wenn es gegenwärtig gebraucht wird, zwangsläufig darin zu finden sein muss. Allerdings können keine Aussagen bezüglich der Zusammensetzung desselben als Korpus getroffen werden, und die Häufigkeit des Auftretens von Lexemen darin muss vorsichtig interpretiert werden. Das Internet kann weder als Querschnitt des privaten Lesestoffs eines Mitglieds der Mittelschicht gesehen werden noch als pauschales Muster der Textproduktion. Bisher ist die Anzahl der Transkripte gesprochener Sprache darin sehr begrenzt, und die geschriebene Sprache ist eine Reflexion der Texte, welche verschiedene Personen ins Internet stellen. Einige dieser Texte existieren nur im Internet,

⁶¹⁵ Vgl. Teubert/ Čermáková (2007): op. cit., S. 77.

⁶¹⁶ Vgl. McEnery/ Hardie (2012): op. cit., S. 8.

⁶¹⁷ Vgl. Teubert/ Čermáková (2007): op. cit., S. 77.

⁶¹⁸ Vgl. Scherer (2006): op. cit., S. 74-75.

andere sind von verschiedenen schriftsprachlichen Materialien übernommen. Bei letzteren raten Teubert/Čermáková allerdings ebenso zur Vorsicht, da nicht alle dieser übernommenen Texte dem Original entsprechen.⁶¹⁹ Zudem ist zu erwähnen, dass das Internet im Gegensatz zu den meisten Korpora eine Mischung aus sorgfältig vorbereiteten Texten und solchen, die nur zufällig zustande gekommen sind, darstellt.⁶²⁰ Lemnitzer/Zinsmeister kritisieren darüber hinaus, dass im Internet in seiner aktuellen Form kaum Metadaten, also Daten über die Herkunft der Texte, deren Entstehungszeitpunkt oder Autorschaft zu finden sind. Diese Situation verbessert sich zwar zunehmend, doch bisher sind die verfügbaren Daten noch nicht ausreichend.⁶²¹

Dennoch ist Scherer der Meinung, das Internet sollte aufgrund der Tatsache, dass es eine immense Menge an realen Sprachdaten enthält, durchaus unter korpuslinguistischen Gesichtspunkten verwendet werden.⁶²² Zudem kann die Mehrsprachigkeit des Internets ihrer Ansicht nach im Fremdsprachenunterricht oder bei Übersetzungen gewinnbringend genutzt werden. Durch Frequenzvergleiche lässt sich auch im Internet herausfinden, welche synonymen Bezeichnungen von Muttersprachlern besonders häufig gebraucht werden, wie zum Beispiel, ob im Englischen vor allem *child*, *infant* oder *kid* für ‚Kind‘ verwendet wird.⁶²³

Dass die Frage „Ist das Internet ein Korpus?“ Linguisten beschäftigt und nicht abschließend geklärt werden kann, sieht man auch an der Veröffentlichung zweier Sammelbände zu diesem Thema: Baroni/Bernardini (2006)⁶²⁴ sowie Hundt et al. (2007)⁶²⁵. Unter den Forschern herrscht also keine Einigkeit in Bezug auf die Frage, ob das Internet ein Korpus ist. Dennoch stimmen Scherer, Kilgarriff/Grefenstette, McEnery/Hardie und Teubert/Čermáková darin überein, dass das Internet durchaus als Quelle für Texte zur Erstellung eines eigenen Korpus genutzt werden kann. Hierfür sollte man alle Texte, mit denen man arbeiten möchte, herunterladen und mit diesen ein

⁶¹⁹ Vgl. Teubert/Čermáková (2007): op. cit., S. 76-77.

⁶²⁰ Vgl. McEnery/Hardie (2012): op. cit., S. 7-8.

⁶²¹ Vgl. Lemnitzer/Zinsmeister (2010): op. cit., S. 43.

⁶²² Vgl. Scherer (2006): op. cit., S. 75.

⁶²³ Vgl. Scherer (2006): op. cit., S. 76.

⁶²⁴ Vgl. Baroni, Marco/ Bernardini, Silvia (Hrsg.) (2006): *WaCky! Working papers on the web as corpus*. Bologna: Gedit.

⁶²⁵ Vgl. Hundt, Marianne/ Nesselhauf, Nadja/ Biewer, Carolin (2007): *Corpus Linguistics and the Web*. Amsterdam and New York: Rodopi.

2. Kategorisierung der Korpora

Spezialkorpus zusammenzustellen. Außerdem ist es ratsam, die Webadressen, zusätzliche bibliographische Informationen sowie das Datum des Downloads zu dokumentieren.⁶²⁶

Kilgariff hat zusammen mit Baroni (2006) bereits zwei Korpora mit jeweils 1,5 Milliarden Lexemen erstellt, das deutsche ‚DeWaC‘ und das italienische ‚ItWaC‘ Korpus, wobei ‚WaC‘ für *Web as Corpus* steht. Er bedauert jedoch auch, dass die Datenmengen der Sprachwissenschaftler nie so groß sein werden wie die von *Google*, *Microsoft* und *Yahoo*. Die Forscher sollten ihre entwickelten Korpora daher allen Interessierten zugänglich machen, um so die Möglichkeiten der Suche ohne kommerzielle Suchmaschinen zu erweitern. Er ist der Meinung, dass das Internet durchaus genutzt werden kann, ohne die gängigen Suchmaschinen zu verwenden.⁶²⁷

Außerdem führen Kilgariff/Grefenstette einige Projekte wie das *EU MEANING project* (Rigau et al., 2002) an, welches das Internet nutzt, um Bedeutungen von Lexemen zu disambiguieren.⁶²⁸ Sie kommen zu dem Ergebnis, dass nicht nur viele sprachwissenschaftliche Untersuchungen, sondern auch die Funktionsweise von *Machine Learning Algorithms* verbessert werden, je mehr Daten vorhanden sind. Die Menge der Daten ist hierbei ihrer Meinung nach wichtiger als verbesserte Algorithmen.⁶²⁹

Wie Kilgariff/Grefenstette bereits 2003 schrieben, eignet sich das Internet auch hervorragend zur schnellen Rechtschreibprüfung: So ergibt die Suche nach *psicología* mittels der Suchmaschine *Google* beispielsweise 62 Mio. Treffer, die Suche nach **psichología* nur 1.080, wodurch sofort klar wird, dass die erste Variante korrekt ist. Zudem stellt es häufig die einzige Quelle für Daten einer bestimmten Sprache dar oder es wird bevorzugt verwendet, weil es kostenlos und immer verfügbar ist.⁶³⁰

Des Weiteren kann der Nutzer die Verbreitung einer Kombination von Lexemen (beispielsweise ein zum Substantiv passendes Adjektiv oder eine Verbindung zwischen Prädikat und Objekt) erörtern, indem er die jeweiligen Ausdrücke in eine Suchmaschine wie *Google* eingibt und überprüft, wie oft diese Kombination zu finden ist. Bei Unsicherheiten bezüglich der Verwendung von Präpositionen wie *por* und *para* oder *de* und *a*, ist es möglich, das fragliche Lexem, beispielsweise ein Verb, mit beiden Präpositionen einzugeben und dabei zu ermitteln, welche Kombination häufiger zu finden

⁶²⁶ Vgl. Teubert/ Čermáková (2007): op. cit., S. 77.

⁶²⁷ Vgl. Kilgariff (2006): op. cit., S. 151.

⁶²⁸ Vgl. Kilgariff/ Grefenstette (2003): op. cit., S. 3-4.

⁶²⁹ Vgl. Kilgariff/ Grefenstette (2003): op. cit., S. 4.

⁶³⁰ Vgl. Kilgariff/ Grefenstette (2003): op. cit., S. 1.

ist. Hierbei muss jedoch die Umgebung untersucht werden, um festzustellen, ob die jeweilige Kombination in einem ähnlichen sprachlichen Kontext wie dem eigenen vorkommt. Dazu sollte der gesuchte Ausdruck in Anführungszeichen geschrieben werden, wie bei „*gracias por su ayuda*“, um auch wirklich nur Ergebnisse mit dem gesamten Suchausdruck und nicht auch Treffer mit nur einem der gesuchten Lexeme zu erhalten. Die Suche mit Suchmaschinen ist noch effektiver, genauer und verlässlicher, wenn der Nutzer die Suchmaschine so einstellt, dass nur spanische Seiten angezeigt werden. Allerdings kann auch hier keine Garantie gegeben werden.

Ein guter Mittelweg zwischen der Suche in Korpora und derjenigen im *World Wide Web* ist das online frei zugängliche Programm *WebCorp Live* der *Birmingham City University*.⁶³¹ Es handelt sich dabei um eine Art ‚Werkzeugsammlung‘, anhand derer das Internet als Korpus durchsucht werden kann. Der Grund für die Entwicklung von *WebCorp Live* war, wie bereits in diesem Kapitel gesehen werden konnte, die Tatsache, dass es immer Aspekte der Sprache geben wird, die zu neu oder zu selten sind, um sie in elektronischen Korpora, egal welcher Größe und Aktualität, finden zu können. Obwohl *WebCorp Live* speziell für die Datensuche im linguistischen Bereich entwickelt wurde, kann es von jedem genutzt werden, der Lexeme und Phrasen untersuchen will, speziell solche, die eben (noch) nicht in Wörterbüchern oder Standardkorpora erscheinen.⁶³² Dabei arbeitet das Programm ‚über‘ den gewöhnlichen Suchmaschinen, die man ausgesucht hat. Es nutzt die Liste der Internetadressen, die von der Suchmaschine ermittelt werden, extrahiert die entsprechenden Konkordanzen und präsentiert diese auf einer einzigen Seite, wobei zusätzlich die Links angegeben werden, um auf die jeweiligen Internetseiten gelangen zu können. Während gewöhnliche Suchmaschinen eine Liste der für den Nutzer relevantesten Internetseiten mit einer Beschreibung der Seite oder einem Ausschnitt daraus als Ergebnis präsentieren und der Nutzer auf jeden Link klicken muss, um die Seite zu sehen, besucht *WebCorp Live* jede einzelne Seite tatsächlich und extrahiert die gesuchten Konkordanzen. Manche Suchmaschinen, wie *Google*, geben zwar eine Art *Key Word in Context* Aufstellung, jedoch auch nicht bei jeder Internetseite,

⁶³¹ Vgl. Research and Development Unit for English Studies, Birmingham City University (ed.): *WebCorp Live. Concordance the web in real-time*. Online Publikation: <http://www.webcorp.org.uk/live/index.jsp> (07.12.2016).

⁶³² Vgl. Research and Development Unit for English Studies, Birmingham City University (ed.): *WebCorp Live. Concordance the web in real-time. Guide*. Online Publikation: <http://www.webcorp.org.uk/live/guide.jsp> (08.12.2016).

2. Kategorisierung der Korpora

und es ist auch nicht erkennbar, wenn mehr als eine Konkordanz zum Suchbegriff auf einer Seite zu finden ist. Für Sprachwissenschaftler ist *WebCorp Live* nicht zuletzt auch wegen weiterer Optionen wie der Auswahl der Spanne zwischen zwei Lexemen bei Konkordanzen oder dem Format, in dem die Ergebnisse präsentiert werden, vorzuziehen. *WebCorp Live* ist durch diese aufwändige Vorgehensweise allerdings langsamer als normale Suchmaschinen. Wenn man sich jedoch vorstellt, man müsste zunächst die jeweiligen Seiten herausfiltern, jede einzeln besuchen und die Konkordanzen in eine Liste kopieren, ist es ein vergleichbar schnelles und zeitsparendes Programm.⁶³³ Für die Suche in *WebCorp Live* wird folgende Benutzeroberfläche verwendet:

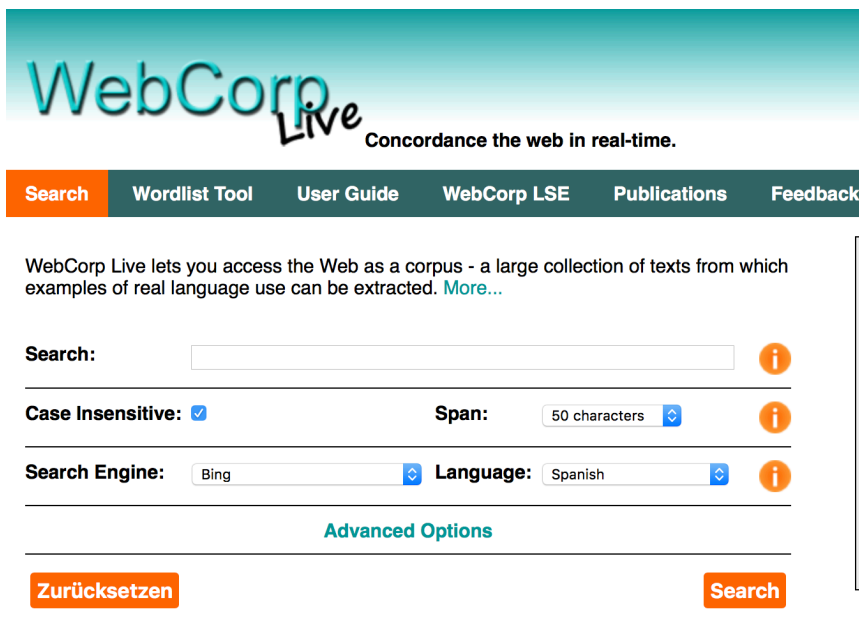
The screenshot shows the WebCorp Live web interface. At the top is a header with the logo 'WebCorp Live' and the tagline 'Concordance the web in real-time.' Below this is a navigation bar with links: Search, Wordlist Tool, User Guide, WebCorp LSE, Publications, and Feedback. The main content area contains a search form. It starts with a 'Search:' label and a text input field. Below this are three rows of options: 'Case Insensitive:' with a checked checkbox, 'Span:' with a dropdown menu set to '50 characters', and 'Search Engine:' with a dropdown menu set to 'Bing'. To the right of each option is an information icon (i). Below these options is a link for 'Advanced Options'. At the bottom of the form are two buttons: 'Zurücksetzen' (Reset) and 'Search'.

Abbildung 74: *WebCorp Live* – Benutzeroberfläche in der Grundeinstellung.⁶³⁴

Der Suchbegriff wird wie gewöhnlich in die obere Leiste eingegeben, wobei keine Anführungszeichen verwendet werden müssen. Zudem kann zwischen Groß- und Kleinschreibung unterschieden werden (indem der Haken bei *Case Insensitive* herausgenommen wird), die Spanne zwischen zwei Lexemen kann ausgewählt werden sowie die zu benutzende Suchmaschine und die gesuchte Sprache. Für das Spanische ist derzeit allerdings nur die Suchmaschine *Bing* verfügbar. In den *Advanced Options* kann

⁶³³ Vgl. Research and Development Unit for English Studies, Birmingham City University (ed.): *WebCorp Live. Concordance the web in real-time. Guide – How does it work?*. Online Publikation: <http://www.webcorp.org.uk/live/guide-how.jsp> (08.12.2016).

⁶³⁴ Research and Development Unit for English Studies, Birmingham City University (ed.): *WebCorp Live. Concordance the web in real-time. Live*. Online Publikation: <http://www.webcorp.org.uk/live/> (29.04.2017).

zusätzlich angegeben werden, ob die Links sowie zusätzliche Informationen zu den Internetseiten angezeigt werden oder nur die Konkordanzen. Zudem kann bestimmt werden, dass nur eine Konkordanz pro gefundener Internetseite angezeigt werden soll, um eine Dominanz der Ergebnisse einer Webseite zu verhindern. Außerdem kann angegeben werden, dass nur auf bestimmten Webseiten gesucht werden soll, und es können spezielle Lexeme für die Suche angegeben oder von der Suche ausgeschlossen werden, um gewisse Ergebnisse bereits im Vorhinein zu vermeiden. Um Lexeme auszugrenzen, muss lediglich ein Minuszeichen vor das Lexem geschrieben werden. Die angegebenen Lexeme müssen dabei nicht zwangsläufig in der Konkordanz erscheinen, sondern lediglich auf der Internetseite zu finden sein.⁶³⁵

Nach der Suche in *WebCorp Live* gibt es verschiedene Möglichkeiten, die Ergebnisse zu analysieren: Es können Kollokate gestaffelt nach den vier Lexemen links und rechts des Suchausdrucks angezeigt werden, mittels der Option *Exclude Stopwords* können besonders häufige Lexeme von der Liste der Kollokate ausgeschlossen werden und die Konkordanzen können alphabetisch oder nach Datum sortiert werden.⁶³⁶

Ein Beispiel für die Nutzung von *WebCorp Live* ist die Suche nach dem relativ neuen Lexem *app*. Sucht man das Lexem *app* im Korpus CREA (siehe Kapitel 2.1.1.1.), so erhält man nur zehn Beispiele aus zwei Dokumenten. Unter diesen 10 Konkordanzen ist jedoch nur einmal die Bedeutung eines Programms für Smartphones zu finden, in den übrigen neun Fällen ist es eine Abkürzung für *aproximadamente*. Im CORPES (siehe Kapitel 2.1.1.2.) finden sich zwar 50 Fälle in 31 Dokumenten, doch auch hier sind andere Bedeutungen zu finden. Das Programm *WebCorp Live* hingegen liefert 189 Beispiele, die allesamt ein Programm für Smartphones bezeichnen. Lässt man sich nun die Kollokationen (siehe Kapitel 1.4.8. und 4.1.1.) anzeigen, so findet man heraus, dass das aus dem Englischen kommende Lexem *app* im Spanischen als feminines Substantiv verwendet wird (der vierte Eintrag auf der Frequenzliste ist der Begleiter *la* mit den

⁶³⁵ Vgl. Research and Development Unit for English Studies, Birmingham City University (ed.): *WebCorp Live. Concordance the web in real-time. Live. Index*. Online Publikation: <http://www.webcorp.org.uk/live/index.jsp> (08.12.2016).

⁶³⁶ Vgl. Research and Development Unit for English Studies, Birmingham City University (ed.): *WebCorp Live. Concordance the web in real-time. Guide - Post*. Online Publikation: <http://www.webcorp.org.uk/live/guide-post.jsp> (08.12.2016).

2. Kategorisierung der Korpora

meisten Treffern als direkt vorausgehendes Lexem). Zudem finden sich Lexeme wie *web*, *móvil*, *google*, *descarga*, *itunes*, *desarrollador* oder *mobile* in der Frequenzliste.⁶³⁷

Derzeit wird eine neue, speziell auf die Bedürfnisse von Linguisten abgestimmte Suchmaschine von der *Birmingham City University* entwickelt, die *WebCorp Linguist's Search Engine (WebCorp LSE)*. Sie wurde entworfen, weil kommerzielle Suchmaschinen für linguistische Studien im Internet nicht geeignet sind, das *WebCorp Live* diese jedoch nutzt, um Ergebnisse aus dem Internet zu erhalten und diese Information für linguistische Zwecke zu ordnen. Das *WebCorp LSE* hingegen wird von einer eigens an der *Birmingham City University* entwickelten Suchmaschine durchsucht. Die speziell entworfenen Programme, wie *Webcrawler*, *Parser*, *Tokeniser* und *Indexer*, erlauben es, große Bereiche des Internets zu durchforsten. Die neue Architektur hat es ermöglicht, nicht nur die Satzerkennung, sondern auch die Datumserkennung, die Ermittlung von Kollokationen und weitere statistische Analyseprogramme, die auch im *WebCorp Live* verfügbar sind, weiterzuentwickeln. Die zusätzliche Vorverarbeitung beinhaltet grammatisches *tagging* und Spracherkennung. Die Benutzeroberfläche erlaubt Suchen nach Lexemen und Phrasen sowie *Wildcards* und Mustererkennung. Derzeit wird das *WebCorp LSE* noch von der Forschungs- und Entwicklungsabteilung für Englische Sprache an der *School of English* in der *Birmingham City University* getestet und ist leider noch nicht online verfügbar.⁶³⁸

Ein weiterer Webservice, der bei Übersetzungen gute Dienste leistet, ist *Linguee*, da hier ein Onlinewörterbuch mit einer Suchmaschine kombiniert wird. *Linguee* bietet Übersetzungen in 25 Sprachen, welche miteinander kombiniert werden können. Dabei wird im Internet nach Texten mit dem eingegebenen Ausdruck gesucht und es werden dem Nutzer parallele Satzpaare angeboten, welche den Ausdruck in der Ausgangssprache und der Zielsprache enthalten.⁶³⁹ *Linguee* verwendet hierfür einen speziellen

⁶³⁷ Vgl. Research and Development Unit for English Studies, Birmingham City University (ed.): *WebCorp Live. Concordance the web in real-time. Suche nach ,app' mittels der Suchmaschine Bing*. Online Publikation: <http://www.webcorp.org.uk/live/cache.jsp?id=FNYVWM62db&info=on&span=c50&colls=on&sort=none&sortpos=1&case=on&date=x> (08.12.2016).

⁶³⁸ Vgl. Research and Development Unit for English Studies, Birmingham City University (ed.): *WebCorp Linguist's Search Engine*. Online Publikation: <http://wse1.webcorp.org.uk/home/guide.html> (02.09.2018).

⁶³⁹ Vgl. Linguee GmbH (ed.): *Linguee. Wörterbuch Englisch-Deutsch und Suche in einer Milliarde Übersetzungen*. Online Publikation: <http://www.linguee.de> (14.11.2016).

Webcrawler, der das Internet nach zweisprachigen Texten durchsucht und diese in parallele Sätze zerlegt. Ein von Menschen trainierter *Machine-Learning-Algorithm* bestimmt dabei die Qualität der Übersetzung. Neben zweisprachigen Texten aus dem Internet, wie Firmenwebsites und Onlineshops, werden Übersetzungen von Patenttexten und EU-Webseiten als Quellen verwendet.⁶⁴⁰

Somit lässt sich festhalten, dass, obgleich sich die Korpuslinguisten nicht darüber einig sind, ob das Internet nun als Korpus gesehen werden kann oder nicht, die Möglichkeiten, die es uns bietet, nicht außer Acht gelassen werden sollten. Nicht zuletzt mit den eben vorgestellten Programmen lässt sich das Internet ähnlich einem Korpus nutzen und ermöglicht damit das Forschen in einer unerschöpflichen Menge an Daten. Im Internet an sich kann zwar nicht nach Wortarten gesucht werden, es können aber bestimmte Ausdrücke, Redewendungen und Texte, in welchen dieselben vorkommen, gefunden werden. Allerdings verfügt es über keine Annotationen, weshalb es für linguistische Untersuchungen nur bedingt verwendet werden kann. Da das Internet nach der für die vorliegende Arbeit geltenden Definition nicht als Korpus kategorisiert werden konnte, wird dessen Nutzbarkeit im Folgenden nicht weiter untersucht werden. Dennoch ist das Internet als Textsammlung durchaus zulässig, sofern der Nutzer im richtigen Umgang mit dem Internet geübt ist.

Wie in diesem Kapitel zu sehen war, gibt es für die spanische Sprache einige sehr gute und interessante Korpora, die in verschiedenen Bereichen und auf unterschiedliche Art und Weise genutzt werden können. Laut Klöden (2002) ist das Spektrum der Korpora durchaus noch erweiterbar. So fehlen ihrer Meinung nach Korpora zur Jugendsprache, zur Medien- und Werbesprache sowie zu Fachsprachen. Zudem sollten in mündlichen Korpora nicht nur *face-to-face*-Kommunikationen erfasst werden, sondern auch Telefonate, Radio- oder Fernsehsendungen, wie beispielsweise Talk-Shows, bei denen die Sprecher sich nicht beobachtet fühlen.⁶⁴¹

Die Erstellung bestimmter Korpora wäre sehr wünschenswert, wie beispielsweise eines spanisch-deutschen Parallelkorpus oder eines umfassenden Lernerkorpus mit deutschen Muttersprachlern und einer spanischen Vergleichsgruppe. Auch der Zugriff

⁶⁴⁰ Vgl. Golem (ed.): *Linguee: Suchmaschine für Übersetzungen*. Online Publikation: <http://www.golem.de/0904/66396.html> (14.11.2016).

⁶⁴¹ Vgl. Klöden (2002): op. cit., S. 13-14.

2. Kategorisierung der Korpora

auf die vollständigen Texte der Referenzkorpora der *Real Academia Española* (siehe Kapitel 2.1.1.1., 2.1.1.2., 2.2.1.1. und 2.2.1.2.) wäre äußerst begrüßenswert. Dennoch verfügt die spanische Sprache bereits über eine Vielzahl nutzbarer Korpora, was nicht für alle Sprachen zutrifft. Mit dem Korpus COLA (siehe Kapitel 2.1.2.1.) ist nun sogar ein Korpus der Jugendsprache verfügbar, was Klöden im Jahr 2002 noch bemängelt hatte.

Wie die einzelnen Korpora nun im fremdsprachlichen Unterricht genutzt werden können, wird im vierten Kapitel behandelt werden. Bevor jedoch genauer auf die Nutzung der Korpora in der Unterrichtspraxis eingegangen wird, soll im folgenden Kapitel auf die dafür nötigen theoretischen Bedingungen sowie auf Motive und Möglichkeiten für den Einsatz von Korpora im Unterricht eingegangen werden.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

Nachdem nun geklärt ist, was unter Korpora und der Korpuslinguistik zu verstehen ist sowie welche Korpora es im Bereich der spanischen Sprache gibt, stellt sich die Frage, ob sich die Korpusarbeit mit dem Fremdsprachenunterricht vereinen lässt. Wie an den Sammelbänden *Corpora and Language Learners* von Aston/Bernardini/Stewart (2004)⁶⁴², Sinclairs *How to use corpora in language teaching* (2004)⁶⁴³ oder *Rethinking Language Pedagogy from a Corpus Perspective* von Burnard/McEnery (2000)⁶⁴⁴ gesehen werden kann, ist die Nutzung von Korpora für den Sprachunterricht im englischsprachigen Bereich bereits seit einiger Zeit weit verbreitet. In diesen Werken finden sich Informationen über und Studien zur Nutzung von Korpora für die Ermittlung dessen, was unterrichtet werden soll, für den Gebrauch im Unterricht an sich sowie zum Vergleich der Sprachkenntnisse von Fremdsprachenlernern und Muttersprachlern. Es wird sowohl Englisch als Muttersprache als auch als Zweit- oder Fremdsprache untersucht sowie deren unterschiedlicher Lernprozess verglichen (vgl. Sripicharn 2004).⁶⁴⁵ Auch verschiedene Stadien des Spracherwerbs werden analysiert, so zum Beispiel bestimmte Bereiche der Grammatik im Englischen von fortgeschrittenen Lernern bei Leńko-Szymańska.⁶⁴⁶ Zudem werden in einigen Studien Lernerkorpora erstellt, um den Sprachstand der Lerner zu ermitteln (vgl. Tankó 2004)⁶⁴⁷, den

⁶⁴² Vgl. Aston, Guy/ Bernardini, Silvia/ Stewart, Dominic (Hrsg.) (2004): *Corpora and Language Learners*. Amsterdam [u.a.]: Benjamins.

⁶⁴³ Vgl. Sinclair, John McHardy (Hrsg.) (2004): *How to Use Corpora in Language Teaching*. Amsterdam [u.a.]: Benjamins.

⁶⁴⁴ Vgl. Burnard, Lou/ McEnery, Tony (Hrsg.) (2000): *Rethinking language pedagogy from a corpus perspective: papers from the Third International Conference on Teaching and Language Corpora*. Frankfurt am Main: Lang.

⁶⁴⁵ Vgl. Sripicharn, Passapong (2004): „Examining native speakers’ and learners’ investigation of the same concordance data and its implications for classroom concordancing with ELF learners“. In: Aston, G./ Bernardini, S./ Stewart, D. (Hrsg.) (2004): *Corpora and Language Learners*. Amsterdam [u.a.]: Benjamins, S. 233-244.

⁶⁴⁶ Vgl. Leńko-Szymańska, Agnieszka (2004): „Demonstratives as anaphora markers in advanced learners’ English“. In: Aston, G./ Bernardini, S./ Stewart, D. (Hrsg.) (2004): *Corpora and Language Learners*. Amsterdam [u.a.]: Benjamins, S. 89-108.

⁶⁴⁷ Vgl. Tankó, Guyla (2004): „The use of adverbial connectors in Hungarian university students’ argumentative essays“. In: Sinclair, J. McH. (Hrsg.) (2004): *How to Use Corpora in Language Teaching*. Amsterdam [u.a.]: Benjamins, S. 157-184.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

Spracherwerb zu analysieren (vgl. Tono 2000)⁶⁴⁸ sowie Fehler der Lerner ausfindig zu machen (vgl. Flowerdew 2000)⁶⁴⁹. Lernerkorpora werden zudem genutzt, um den Sprachunterricht zu verbessern, so wie es beispielsweise im Aufsatz “How learner corpus analysis can contribute to language teaching: A study of support verb constructions” von Nesselhauf beschrieben wird.⁶⁵⁰

Im spanischsprachigen Bereich gibt es leider nur wenige Studien zur Nutzung von Korpora, was auch Parodi in der Einleitung zu *Working with Spanish Corpora* (2007)⁶⁵¹ bedauert. Sowohl in englischsprachigen als auch in spanischsprachigen Ländern seien Studien zur spanischen Korpuslinguistik selten, trotz der zunehmenden Bedeutung des Spanischen als internationaler Sprache, der wachsenden Anzahl an Sprechern und der Tatsache, dass es sozusagen die zweite *lingua franca* neben dem Englischen ist.⁶⁵² Parodis Sammlung enthält allerdings hauptsächlich allgemeine Studien zur Nutzung von Korpora für die Weiterentwicklung der Beschreibung der spanischen Sprache insbesondere im Hinblick auf Variation in verschiedenen Registern und lediglich eine Studie zur Nutzung eines Lernerkorpus sowie zwei Studien, deren Erkenntnisse trotz allgemeiner Zielsetzungen für den Fremdsprachenunterricht verwendet werden können.

Parodi selbst untersuchte die Variation der spanischen Sprache in geschriebenen und gesprochenen sowie spezialisierten und nicht-spezialisierten Registern anhand des Korpus PUCV2003, einem Korpus mit annähernd 2,5 Mio. Lexemen.⁶⁵³ Ciapuscio analysierte gesprochene akademische Diskursgenres anhand des COTECA Korpus (*Corpus Textual del Español Científico de la Argentina*), eines argentinischen Korpus wissenschaftlicher Texte, um mehr über das gesprochene und geschriebene Spanisch

⁶⁴⁸ Vgl. Tono, Yukio (2000): „A computer learner corpus based analysis of the acquisition order of English grammatical morphemes“. In: Burnard, L./ McEnery, T. (Hrsg.) (2000): *Rethinking language pedagogy from a corpus perspective: papers from the Third International Conference on Teaching and Language Corpora*. Frankfurt am Main: Lang, S. 123-132.

⁶⁴⁹ Vgl. Flowerdew, Lynn (2000): „Investigating referential and pragmatic errors in a learner corpus“. In: Burnard, L./ McEnery, T. (Hrsg.) (2000): *Rethinking language pedagogy from a corpus perspective: papers from the Third International Conference on Teaching and Language Corpora*. Frankfurt am Main: Lang, S. 145-154.

⁶⁵⁰ Vgl. Nesselhauf, Nadja (2004): „How learner corpus analysis can contribute to language teaching: A study of support verb constructions“. In: Aston, G./ Bernardini, S./ Stewart, D. (Hrsg.) (2004): *Corpora and Language Learners*. Amsterdam [u.a.]: Benjamins, S. 109-124.

⁶⁵¹ Vgl. Parodi, Giovanni (Hrsg.) (2007): *Working With Spanish Corpora*. London: Continuum.

⁶⁵² Vgl. Parodi (2007a): op. cit., S. 1.

⁶⁵³ Vgl. Parodi, Giovanni (2007b): „Variation across registers in Spanish: Exploring the El Grial PUCV Corpus“. In: Parodi, G. (Hrsg.) (2007): *Working With Spanish Corpora*. London: Continuum, S. 11-53.

argentinischer Wissenschaftler und Spezialisten zu erfahren.⁶⁵⁴ Sedano verglich die Verwendung des morphologischen und des periphrastischen Futurs im Hinblick auf zeitliche Distanz, grammatische Person und den Wahrheitsgehalt einer Aussage. Ihre Ergebnisse sind auch für den Spanischunterricht durchaus zu berücksichtigen, wie in Kapitel 3.3.2.2. zu sehen sein wird.⁶⁵⁵ Parodi/Gramajo veranschaulichten, wie eine korpusbasierte Studie zur Charakterisierung eines speziellen akademisch-beruflichen Diskursgenres dazu dienen kann, nicht nur eine Sammlung spezialisierter Texte aus verschiedenen Perspektiven zu beschreiben, sondern auch eine Texttypologie mit didaktischer Zielsetzung zu verfeinern.⁶⁵⁶ López-Ferrero verglich den Gebrauch parataktischer Konjunktionen in einem Lernerkorpus mit demjenigen in einem Spezialkorpus, um die syntaktischen, semantischen, pragmatischen und Diskurseigenschaften von parataktischen Konjunktionen wie *y/e*, *ni*, *o/u*, *bien ... bien* oder *pero* genau beschreiben zu können.⁶⁵⁷ Des Weiteren untersuchten Tracy-Ventura/Cortes/Biber lexikalische Bündel, die gewöhnlichsten sich wiederholenden Wortsequenzen in einer Sprache, in gesprochenen und geschriebenen Texten.⁶⁵⁸

Außerdem gibt es die Aufsatzsammlung *Lingüística de corpus y adquisición de la lengua* von Fernández Pérez, welche sich jedoch mit dem Erstspracherwerb spanischer Kinder und nicht mit dem Fremdspracherwerb beschäftigt.⁶⁵⁹

Das zuvor erwähnte *Corpora and Language Learners* umfasst zudem einen Artikel von Mark Davies, dem Ersteller mehrerer großer Korpora wie des CdE (siehe Kapitel 2.1.1.3., 2.2.1.3. und 2.6.1.) und des *Corpus of Contemporary American English* (COCA), mit dem Thema „Student use of large annotated corpora to analyze syntactic variation“. In seiner Studie beschreibt er, wie fortgeschrittene Lerner des Spanischen

⁶⁵⁴ Vgl. Ciapuscio, Guiomar (2007): „Epistemic modality and academic orality: Pilot study for COTECA (Corpus Textual del Español Científico de la Argentina)“. In: Parodi, G. (Hrsg.) (2007): *Working With Spanish Corpora*. London: Continuum, S. 90-105.

⁶⁵⁵ Vgl. Sedano, Mercedes (2007): „Future tense expressions in several Spanish corpora“. In: Parodi, G. (Hrsg.) (2007): *Working With Spanish Corpora*. London: Continuum, S. 132-144.

⁶⁵⁶ Vgl. Parodi, Giovanni/ Gramajo, Aída (2007): „Technical-professional discourses: Specialized and dissemination text types“. In: Parodi, G. (Hrsg.) (2007): *Working With Spanish Corpora*. London: Continuum, S. 145-172.

⁶⁵⁷ Vgl. López-Ferrero, Carmen (2007): „Academic writing: Exploring *Corpus 92*“. In: Parodi, G. (Hrsg.) (2007): *Working With Spanish Corpora*. London: Continuum, S. 173-194.

⁶⁵⁸ Vgl. Tracy-Ventura/ Cortes/ Biber (2007): op. cit., S. 217-231.

⁶⁵⁹ Vgl. Fernández Pérez, Milagros (2011): *Lingüística de corpus y adquisición de la lengua*. Madrid: Arco Libros.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

verschiedene große Sprachkorpora nutzten, um bestimmte syntaktische Phänomene in der spanischen Sprache zu untersuchen.⁶⁶⁰ Außerdem veröffentlichte er eine Studie in der Sammlung *Rethinking Language Pedagogy from a Corpus Perspective*, in welcher er computerbasierte Korpora für fortgeschrittene Kurse in spanischer Linguistik an der *Illinois State University* verwendete. Er nutzte die Korpora beispielsweise, um den Studenten in den Kursen *Approaches to Spanish Syntax* und *History of the Spanish Language* Veränderungen in der Syntax während der letzten 800 Jahre sowie syntaktische Variation in verschiedenen Dialekten und Registern des aktuellen Spanisch vor Augen zu führen.⁶⁶¹

Zudem gibt es weitere Studien, welche zwar nicht eigens für den Fremdsprachenunterricht durchgeführt wurden, deren Erkenntnisse für diesen allerdings durchaus genutzt werden können. Es wird jedoch weder bei den bisher beschriebenen noch bei den folgenden Studien Anspruch auf Vollständigkeit erhoben.

Alcántara Plá untersuchte linguistische Strukturen anhand eines Korpus und versuchte, in seinem einführenden Werk *Introducción al análisis de estructuras lingüísticas en corpus* zu zeigen, wie Texte auf einfache und gewinnbringende Weise annotiert werden können. Dabei erstellte er exemplarisch eine kurze Grammatik der spontanen gesprochenen Sprache, welche sich aufgrund des semantischen Schwerpunkts von den traditionellen Grammatiken abhebt, allerdings nicht speziell für Lerner des Spanischen verfasst wurde.⁶⁶² Alvarado Ortega nutzte Korpora, um *fórmulas rutinarias* der spanischen Sprache, wie *¡madre mía!* oder *¡vaya tela!*, zu analysieren. Diese gewohnheitsmäßigen Floskeln wurden bereits in einigen Veröffentlichungen unter lexikographischen, formalen und psycho-sozialen Gesichtspunkten thematisiert, allerdings immer nur auf der Basis von Daten geschriebener Sprache.⁶⁶³ Alvarado Ortega

⁶⁶⁰ Vgl. Davies, Mark (2004): „Student use of large, annotated corpora to analyze syntactic variation“. In: Aston, G./ Bernardini, S./ Stewart, D. (2004): *Corpora and Language Learners*. Amsterdam [u.a.]: Benjamins, S. 259-269.

⁶⁶¹ Vgl. Davies, Mark (2000): „Using multi-million word corpora of historical and dialectal Spanish texts to teach advanced courses in Spanish linguistics“. In: Burnard, L./ McEnery, T. (Hrsg.) (2000): *Rethinking language pedagogy from a corpus perspective: papers from the Third International Conference on Teaching and Language Corpora*. Frankfurt am Main: Peter Lang, S. 173-185.

⁶⁶² Vgl. Alcántara Plá (2007): op. cit.

⁶⁶³ Vgl. Alvarado Ortega, María Belén (2008): *Las fórmulas rutinarias en el español actual*. Tesis Doctorales, Universidad de Alicante. Online Publikation: <https://www.google.de/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&cad=rja&uact=8&ved=2ahUKEwjL97GXhoLjAhXNJVAKHZbLAl0QFjABegQIABAC&url=http%3A%2F%2Fwww.cervantesvirtual.c>

betrachtete sie nun aufgrund ihrer Bedeutung für die Konversation in verschiedenen Korpora gesprochener Sprache.⁶⁶⁴ Vande Castele untersuchte die intrinsischen Eigenschaften der verschiedenen appositiven Konstruktionen sowie die kommunikativen Gründe für deren Einsatz im journalistischen, informativen Bereich anhand eines Korpus und verglich sie mit anderen Ausdrücken, die in ähnlichen Kontexten eingesetzt werden.⁶⁶⁵

Im Verlauf der Arbeit, insbesondere in den Kapiteln 3.3.2. und 3.3.3., werden einige der eben erwähnten Studien genauer beschrieben sowie weitere Studien angeführt werden. Es stellt sich jedoch zunächst die Frage, ob Korpora auch im deutschsprachigen Raum in den Unterricht integriert werden können und inwiefern sich die Korpusarbeit mit den Rahmenbedingungen des Fremdsprachen- und insbesondere des Spanischunterrichts vereinbaren lässt.

In diesem Kapitel soll daher zunächst geklärt werden, wie sich die Arbeit mit Korpora hinsichtlich des lerntheoretischen Hintergrunds in die schulische Praxis integrieren lässt. Es werden die Ansätze und Methoden des Fremdsprachenunterrichts im Allgemeinen besprochen, der *Gemeinsame Europäische Referenzrahmen für Sprachen*, welchen es seit dem Jahr 2001 zu berücksichtigen gilt, sowie der Lehrplan für Spanisch am bayerischen Gymnasium als Beispiel für einen der Lehrpläne für deutsche Schüler untersucht. Aus dieser Analyse kristallisieren sich einige für den Fremdsprachenunterricht wichtige Kriterien heraus, welche durch die Nutzung von Korpora besonders gefördert werden können und daher im Anschluss genauer betrachtet werden: das umstrittene Kriterium der Authentizität, die kommunikative Kompetenz sowie die Lernerautonomie. Schließlich wird die wichtigste Frage geklärt, nämlich diejenige danach, wie die Korpora nun für den Unterricht genutzt werden können. Es wird eine Einteilung der verschiedenen Nutzungsmöglichkeiten vorgenommen, welche an gegebener Stelle anhand von Studien aus dem Englischen, Französischen oder Spanischen veranschaulicht wird.

[om%2FdescargaPdf%2Fflas-formulas-rutinarias-en-el-espaol-actual%2F&usg=AOvVaw1HsmoPE083-C6tN5A71l6l](#) (24.06.2019), S. 87-89.

⁶⁶⁴ Vgl. Alvarado Ortega, María Belén (2010): *Las fórmulas rutinarias del español: teoría y aplicaciones*. Frankfurt am Main: Lang.

⁶⁶⁵ Vgl. Vande Castele, An (2010): *Las construcciones apositivas de referente humano en el discurso periodístico informativo*. Tübingen: Narr Francke Attempto.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

3.1. Lerntheoretischer Hintergrund

Im Hinblick auf die Vereinbarkeit der Korpusarbeit mit dem Fremdsprachenunterricht an deutschen Schulen werden nun also zunächst allgemeine Ansätze und Methoden des Fremdsprachenunterrichts, der *Gemeinsame Europäische Referenzrahmen für Sprachen* sowie der Lehrplan für die Fremdsprache Spanisch an bayerischen Gymnasien genauer betrachtet.

3.1.1. Ansätze und Methoden des Fremdsprachenunterrichts

Nach der Grammatik-Übersetzungsmethode mit der ausschließlichen Einübung der Schriftsprache, der direkten Methode mit der Konzentration auf die gesprochene Sprache und der audiolingualen beziehungsweise audiovisuellen Methode mit dem *pattern drill* wurde zu Beginn der 1970er Jahre mit der Veröffentlichung des Buchs „Kommunikative Kompetenz als übergeordnetes Lernziel im Englischunterricht“ von Hans-Eberhard Piepho die *Kommunikative Wende* eingeläutet. Im Unterschied zur bisherigen Betonung von Grammatik und Übersetzung, Nachahmung und Wiederholung wird das Lernen von nun an als ein kognitiver und kommunikativer Prozess gesehen. Der kommunikative Ansatz konzentriert sich nicht wie bis dahin auf die Lehrstoffvermittlung und die Lehrerperspektive, sondern stellt den Lerner in den Mittelpunkt.⁶⁶⁶ Außerdem wird die sprachliche Korrektheit in den Hintergrund gedrängt und anstatt dessen die angemessene sprachliche Kommunikation je nach Situation und Gesprächspartner betont. Laut Fäcke prägt die Orientierung der kommunikativen Didaktik im Hinblick auf die kommunikative Kompetenz in realen Alltagssituationen und die Angemessenheit der Kommunikation an sich den Spanischunterricht bis heute. In zahlreichen Lehrplänen und Rahmenrichtlinien einzelner Bundesländer fänden sich Aussagen, welche die kommunikative Kompetenz an oberste Stelle setzten. Sie legt zudem dar, dass die kommunikative Didaktik die Grundlage des modernen Spanischunterrichts seit den 1970er Jahren bildet, da es danach keine weitere eigenständige Methode gegeben hätte. Die danach anzufindenden

⁶⁶⁶ Vgl. Grünewald, Andreas/ Küster, Lutz (2009): *Fachdidaktik Spanisch. Tradition, Innovation, Praxis*. Stuttgart: Klett, S. 100-103.

Methoden stellten lediglich Weiterentwicklungen dar, welche auf der kommunikativen Didaktik beruhten, ohne sie grundlegend infrage zu stellen.⁶⁶⁷

Laut Grünewald/Küster entwickelte sich nach dem kommunikativen Ansatz ab Mitte der 1980er Jahre der interkulturelle Ansatz. Allerdings geben auch sie zu, dass dieser den kommunikativen Ansatz erweitert und wie jener kein konsistentes Methodenkonzept entwickeln, sondern lediglich neue übergreifende Ziele und Themen definieren soll. Der interkulturelle Ansatz ist aus dem pragmatisch-funktionalen Konzept, dem audiovisuellen und eben dem kommunikativen Ansatz entstanden und setzt es sich zum Ziel, die Schüler im Fremdsprachenunterricht auf den zunehmenden Kontakt mit zielsprachlichen Personen vorzubereiten. Im Zeitalter der Kommunikation mittels moderner Medien, des Zusammenwachsens Europas und der wachsenden privaten wie beruflichen Mobilität soll den Schülern dabei geholfen werden, Vorurteile abzubauen, Toleranz zu fördern und interkulturelle Kontakte aufzubauen. Die Zielkultur und Zielsprache stehen bei diesem Ansatz im Vordergrund, ebenso wie die Verwendung von Unterrichtsmaterialien wie Computer, Bilder, Fotos, Videos, Hörtexte, Musik und Kunst. Fremdsprachliche Texte dienen häufig nicht nur als Basis des Spracherwerbs, sondern auch als Ausgangspunkt für Vergleiche zwischen der Ziel- und der Ausgangskultur. Über die Ermittlung von Gemeinsamkeiten und Unterschieden soll Verständnis für die Zielgruppe, deren Denken, Handeln und Sprache geschaffen werden. Daher soll sich im Unterricht verstärkt mit landeskundlichen und kulturspezifischen Texten beschäftigt werden. Auf diesem Ansatz aufbauend ist die Verwendung ‚realer‘ (vgl. Diskussion des Begriffs ‚Authentizität‘ in Kapitel 3.2.1.), von Muttersprachlern geschriebener Texte zu propagieren, da das Denken und Handeln der zielsprachlichen Kultur an diesen Texten am besten gesehen werden kann. Vergleiche zwischen der fremden und der eigenen Kultur anhand von Texten sind vor allem dann real, wenn sie auf einer realen Grundlage beruhen. Dieser Kontakt mit anderen Ländern, Kulturen und Sprachen führt zunehmend zu interkulturellen Begegnungen, Fremdheitserfahrungen und der Bildung von Vorurteilen, welche es auch im Unterricht durch Förderung von Empathiefähigkeit und Toleranz vorzubereiten beziehungsweise welchen es vorzubeugen gilt.⁶⁶⁸

⁶⁶⁷ Vgl. Fäcke, Christiane (2011): *Fachdidaktik Spanisch. Eine Einführung*. Tübingen: Narr Francke Attempto, S. 43-46.

⁶⁶⁸ Vgl. Grünewald/ Küster (2009): op. cit., S. 103.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

Kramersch gibt an, dass den kommunikativen Ansätzen gemäß, egal ob diese funktional-notional ausgerichtet sind (wie in den 70-er Jahren) oder leistungsorientiert (wie in den 80-er Jahren), die Lerner so viele gesprochene oder geschriebene und nicht eigens für pädagogische Zwecke erstellte Texte wie möglich lesen beziehungsweise hören sollten. Auf diese Weise sollen die Lerner durch die ‚Authentisierung‘ der Kommunikation die Sprechgewohnheiten und Lebensstile des Ziellandes besser verstehen und sich so in muttersprachlicher Umgebung angemessener verhalten.⁶⁶⁹

Leitzke-Ungerer begründet die Nutzung realer Texte im Unterricht wie folgt:

„Die Forderung nach Authentizität ergibt sich aus der grundlegenden Zielsetzung des Fremdsprachenunterrichts: Die Schülerinnen und Schüler sollen sprachliche und interkulturelle Handlungsfähigkeit erwerben, um in der Zielkultur, genauer gesagt, in realen Kommunikationssituationen sprachlich und interkulturell angemessen agieren zu können. Authentizität stellt nach diesem Verständnis die Brücke zur zielsprachlichen und zielkulturellen Wirklichkeit her und wird in diesem Sinn auch in den Curricula der Bundesländer als eines der zentralen Prinzipien des Fremdsprachenunterrichts hervorgehoben.“⁶⁷⁰

Sie wundert sich allerdings darüber, dass dieses Thema zwar in Lehrplänen und Rahmenrichtlinien aufgeführt, aber in einführenden Werken in die Fremdsprachendidaktik nicht diskutiert wird.⁶⁷¹

Fäcke gibt als Gründe für die Verwendung realer Materialien einen unverfälschten Bezug zur Realität spanischsprachiger Länder, aktuelle und echte Vermittlungsperspektiven und zudem sprachliche wie inhaltliche Echtheit an. Außerdem bestehe bei didaktisierten Texten die Gefahr der Verfälschung durch sprachliche und zugleich inhaltliche Glättungen sowie Vereinfachungen und damit eine Entfremdung vom Originaltext.⁶⁷²

Korpora können somit neben dem Internet ein wichtiges Bindeglied zwischen den Fremdsprachenlernern und realen Sprachdaten darstellen. Während die Lerner jedoch im Internet bisweilen auf Texte stoßen, welche nicht von Muttersprachlern geschrieben wurden, enthalten spanischsprachige Korpora ausschließlich Texte von

⁶⁶⁹ Vgl. Kramersch, Claire J. (1993): *Context and culture in language teaching*. Oxford: Oxford Univ. Press, S. 185.

⁶⁷⁰ Leitzke-Ungerer, Eva (2010): „Zielkulturelle und lernkontextbezogene Authentizität im Fremdsprachenunterricht“. In: Frings, M./ Leitzke-Ungerer, E. (Hrsg.) (2010): *Authentizität im Unterricht romanischer Sprachen*. Stuttgart: ibidem-Verlag, S. 11.

⁶⁷¹ Vgl. Leitzke-Ungerer (2010): op. cit., S. 11.

⁶⁷² Vgl. Fäcke (2011): op. cit., S. 220-221.

Muttersprachlern, da sie zuvor von den Korpuserstellern unter bestimmten Gesichtspunkten speziell zur Korpuserstellung ausgewählt wurden.

Gegenwärtig ist der Fremdsprachenunterricht von einer Lerner- und Prozessorientierung geprägt, das heißt, er ist kommunikativ und handlungsorientiert ausgerichtet, die Schüler und ihr individueller Lernprozess stehen im Mittelpunkt. Es sollen effiziente und interessante Lernangebote bereitgestellt werden, damit die Lerner die angestrebten Kompetenzen in der Fremdsprache erlangen können.⁶⁷³ In diesem Zusammenhang steht zudem das Prinzip der Ganzheitlichkeit, das heißt, dass Erfahrungen geschaffen werden sollen, die Intellekt, Gefühl und Sinne ansprechen, was beispielsweise durch abwechslungsreiche Unterrichtsszenarien, die das Interesse der Schüler wecken, erreicht werden kann.⁶⁷⁴ Der Einbau von Unterrichtsphasen, in denen mit einem Korpus gearbeitet wird, kann in diesem Sinne eine abwechslungsreiche Lerneinheit bieten, welche sich stark vom gewöhnlichen Unterricht unterscheidet. Bei technik- und computerinteressierten Schülern, womit in der aktuellen Generation wohl die Mehrheit angesprochen sein dürfte, weckt der Umgang mit dem Computer zusätzlich deren individuelle Interessen.

Nachdem man in den 1990er Jahren, als der Computer Einzug in den Unterricht hielt, glaubte, er könne den Lehrer ersetzen, und Begriffe wie *computer based language learning* fielen, spricht man heute vielmehr vom *computer assisted language learning* (CALL). Der Computer wird als ein Medium unter anderen gesehen. Neben der Verwendung des Computers für Übungsprogramme, Simulationen, Lernsoftware, digitalisierte Lexika und Enzyklopädien, dient er auch als Zugang zum Internet. Das Internet ermöglicht fremdsprachliche ‚Authentizität‘ im Klassenzimmer, da es ein Medium für den Kontakt zu Muttersprachlern ist (E-Mails, Chatrooms, Newsgroups und Foren) und fremdsprachliche Texte, sowohl aus der tagesaktuellen Presse als auch aus der Literatur, bereithält. Zudem sind Computer und Internet ein Teil der Lebenswelt der Schüler und sollten schon allein aus diesem Grund in den Unterricht miteinbezogen werden. Des Weiteren stellt das Internet ein Fenster zur Welt dar, unterstützt

⁶⁷³ Vgl. Grünewald/ Küster (2009): op. cit., S. 108.

⁶⁷⁴ Vgl. Grünewald/ Küster (2009): op. cit., S. 113-114.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

selbstgesteuerte Lernprozesse und fördert Motivation und Medienkompetenz seitens der Schüler.⁶⁷⁵

Ein weiteres wichtiges Prinzip des aktuellen Fremdsprachenunterrichts ist die Handlungsorientierung. Das Lernen soll eine aktive und zunehmend selbstgesteuerte Tätigkeit des Schülers sein, wobei die Lehrkraft den Prozess kompetent begleitet.⁶⁷⁶ Im Sinne der Kognitionsorientierung rückt dabei die Lernerautonomie zunehmend in den Vordergrund. Es wird die bis dahin als Lernerautonomie verstandene selbstverantwortliche Steuerung des eigenen Lernens erweitert um die Autonomie als Neustrukturierung von Lernbedingungen sowie um die Einstellung, sich selbst als verantwortlich für eigene Entscheidungen im Miteinander mit anderen zu sehen.⁶⁷⁷ Im gemäßigten Konstruktivismus wird das Lernen als „selbstgesteuerter, autonomer, selbstverantwortlicher Konstruktionsprozess“ interpretiert.⁶⁷⁸ Wie in diesem Kapitel zu sehen sein wird, ist eines der Ziele der Arbeit mit Korpora, dass der Lerner selbst im Korpus forscht und dort seine eigenen Erfahrungen macht sowie selbstständig die Antworten auf zuvor gestellte Fragen findet. Er soll dabei von der Lehrperson angeleitet und unterstützt werden, welche Hinweise und Tipps gibt, ihn jedoch nicht zu sehr in eine bestimmte Richtung lenkt.

Eines der bedeutendsten aktuellen Unterrichtsprinzipien ist allerdings die Aufgabenorientierung, auch *enfoque por tareas* genannt. Laut ihren Befürwortern vereint sie die Vorzüge der kommunikativen Methode mit einer funktionalisierten Strukturorientierung, was dazu führt, dass die Nachteile des traditionellen Grammatikunterrichts umgangen werden. Es steht also nicht das Erlernen der Schriftsprache anhand von speziell auf Wortschatz und Grammatik zugeschnittenen Übungssätzen unter Vernachlässigung der Sprechfertigkeit im Vordergrund, sondern die Schüler sollen in verschiedensten Situationen ihre Kommunikativität zur Erreichung eines Ziels unter Beweis stellen. Eine *tarea* ist eine Aufgabe, deren Ziel es ist, die Sprache sinnvoll und kreativ einzusetzen, um ein Produkt zu schaffen. Dabei kann es sich laut Grünewald/Küster um die Vorbereitung und Gestaltung eines Posters, eines Musikvideos, einer Statistik oder eines Interviews handeln, also um alles, wozu Kreativität und

⁶⁷⁵ Vgl. Fäcke (2011): op. cit., S. 62-63.

⁶⁷⁶ Vgl. Grünewald/ Küster (2009): op. cit., S. 115-116.

⁶⁷⁷ Vgl. Fäcke (2011): op. cit., S. 55-56.

⁶⁷⁸ Fäcke (2011): op. cit., S. 59.

„authentische“ Sprache benötigt werden (zur Authentizität siehe Kapitel 3.2.1.). Diese Ausrichtung des Unterrichts an Aufgaben entstand aufgrund der Unzufriedenheit erwachsener Lerner über die mangelnde Vorbereitung auf reale Sprechsituationen. Deshalb sind die *tareas* so konzipiert, dass sie den Zweck und das Ergebnis einer Aktivität nennen und nicht auf der Verwendung einer bestimmten Form beruhen, wobei die Sprache möglichst „authentisch“ verwendet werden soll. Obgleich der Erfolg dieses Konzepts gegenüber traditionellen Ansätzen noch nicht belegt ist, herrscht laut Grünewald/Küster Einigkeit darüber, dass die Aufgabenorientierung eine wertvolle methodische Weiterentwicklung des Fremdsprachenunterrichts darstellt.⁶⁷⁹ Die Arbeit mit Korpora kann diesen Ansatz neben der Bereitstellung natürlicher, realer Sprachdaten auch darin unterstützen, dass der Unterricht im Umgang mit dem Korpus lernerzentriert und weniger von der Lehrkraft gesteuert und kontrolliert ist.

Zudem unterscheidet sich eine *tarea* laut Grünewald/Küster von einer Übung darin, dass der Lerner auch sprachlich Neues kennenlernen soll, und zwar dann, wenn er eine Wissenslücke beispielsweise in seinen Grammatikkenntnissen feststellt oder sein Wissen in einem bestimmten Bereich vertiefen will, da dies zur Lösung der Aufgabe notwendig ist. Dabei bietet der Umgang mit dem Korpus allein keine vollständige *tarea* im Sinne des *enfoque por tareas*, da kein Produkt geschaffen wird und die Arbeit mit dem Korpus meist form- und nicht inhaltsorientiert ist. Das Korpus bietet allerdings die Möglichkeit, die Lerner bei der Erreichung des Lernziels zu unterstützen, indem es in einer sogenannten *formorientierten Teilaufgabe* vor oder nach der eigentlichen Lernaufgabe eingesetzt wird, „wenn dem Lerner bewusst ist, dass er z. B. die Kenntnis einer bestimmten Struktur/Lexik zur Lösung der kommunikativen Aufgabe benötigt“.⁶⁸⁰ Wenn die Schüler also während ihrer Arbeit an der *tarea final* auf eine Struktur oder einen Bereich der Lexik stoßen, den sie zur Erreichung des Ziels benötigen, können sie ein Korpus heranziehen, um sich über die Verwendung zu informieren. Ebenso ist es möglich, dass die Lernenden am Ende einer Aufgabe feststellen, dass sie bei der Bearbeitung einige sprachliche Probleme hatten, über welche sie nun mehr erfahren wollen, um ihre Sprachkompetenz zu verbessern. Zu diesem Zeitpunkt haben manche

⁶⁷⁹ Vgl. Grünewald/ Küster (2009): op. cit., S. 100-101, 116-117.

⁶⁸⁰ Grünewald/ Küster (2009): op. cit., S. 118.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

Schüler ihre Defizite selbstständig erkannt, so Grünewald, und verspüren idealerweise ein echtes Lernbedürfnis, diese zu beheben.⁶⁸¹

Auch Alonso Pérez-Ávila geht davon aus, dass die Korpusarbeit im Fremdsprachenunterricht gut mit dem Konzept des *task-based learning* oder *enfoque por tareas* kombiniert werden kann, da es hier immer einen gemeinsamen Nenner gibt, der durch den Lernprozess führt, sei dieser auch noch so vielfältig: die Schüler müssen dazu in der Lage sein, eine *tarea final* auszuführen und alle *tareas intermedias*, also die Aufgaben auf dem Weg dorthin, dienen dazu, den Schülern die Fähigkeiten und Fertigkeiten zu vermitteln, die sie für die *tarea final* benötigen. Aufgrund der großen Freiheit, welche die Schüler in dieser neuen Art des Lernprozesses innerhalb der Arbeit mit Korpora genießen, wird es allerdings schwierig sein, die Inhalte und Ziele, die erreicht werden sollen, zu kontrollieren. Verschiedene Konkordanzen (siehe Kapitel 1.4.3.) können zu weiteren Suchen und Diskussionen über die Ergebnisse führen, was im Allgemeinen positiv zu werten ist, wodurch jedoch auch das Ziel einer Unterrichtseinheit aus den Augen verloren werden kann. Wenn man die Arbeit mit einem Korpus nun allerdings im Sinne der *enseñanza por tareas* sieht, heißt das, so sehr das Korpus, die Informationstechnologie oder die Freiheit, das auszuwählen, was man lernen möchte, den Lerner auch vom Weg abbringen können, es gibt immer eine Referenz, die sowohl den Schülern als auch dem Lehrer dabei hilft, sich nicht in der Fülle an Möglichkeiten zu verlieren, welche die neuen Werkzeuge bieten.⁶⁸² Alonso Pérez-Ávila geht also davon aus, dass die Korpusarbeit an sich auch eine *tarea* darstellen kann. Dies wäre beispielsweise der Fall, wenn die *tarea* darin besteht, ein Märchen zu schreiben, und dazu ein Märchenkorpus herangezogen oder sogar erstellt werden soll, um zu erfahren, wie Märchen für gewöhnlich aufgebaut sind und welche sprachlichen Strukturen darin häufig verwendet werden. Außerdem könnte ein Korpus verwendet werden, um sich eine bestimmte sprachliche Struktur anzueignen, indem diese im Korpus untersucht wird. Auch die Erstellung eines Lernerkorpus wäre im Sinne einer *tarea* denkbar.

⁶⁸¹ Vgl. Grünewald/ Küster (2009): op. cit., S. 120.

⁶⁸² Vgl. Alonso Pérez-Ávila, Elena (2007): *El corpus lingüístico en la didáctica del léxico del español como LE*. Online Publikation:
<https://aperto.unito.it/retrieve/handle/2318/57837/6964/El%20corpus%20lingüístico%20en%20la%20didáctica%20del%20léxico%20del%20español%20como%20LE.pdf> (06.06.2017), S. 10-11.

Ein weiterer Pfeiler des aktuellen Fremdsprachenunterrichts ist die Mehrsprachigkeitsdidaktik, deren Augenmerk auf den kommunikativen Kompetenzen und dem allgemeinen Weltwissen liegt.⁶⁸³ In der Mehrsprachigkeitsdidaktik sollen vor allem Parallelen zwischen romanischen Sprachen genutzt und daraus zumindest passive Sprachkenntnisse gefördert werden. Lerner des Spanischen sollen beispielsweise Texte in anderen romanischen Sprachen lesen können. Die sprachpolitische Zielsetzung besteht darin, dass alle Bürger der EU neben der eigenen Muttersprache in mindestens zwei Fremdsprachen kommunizieren können und dem Englischen als *lingua franca* damit weniger Bedeutung zukommt.⁶⁸⁴ Dennoch wurden die Zielvorstellungen eines mehrsprachigen Europas, so wie es der Europarat im Jahr der Sprachen 2001 gefordert hat, nicht erfüllt, denn das Ziel der Mehrsprachigkeit liegt noch in weiter Ferne. Die Forderung nach kommunikativer Kompetenz, um größtmögliche Mobilität sowie ein besseres Verstehen und Zusammenarbeiten in Europa zu erreichen, ist jedoch aktueller denn je im Fremdsprachenunterricht.⁶⁸⁵

Zusätzlich zu den bisher genannten Ansätzen und Methoden unterscheidet man instruktivistische und konstruktivistische Theorien in der Sprach- und Kulturvermittlung. Beide sehen den Kognitivismus als Teil ihrer Wurzeln, doch sie unterscheiden sich bei der Zuschreibung der hauptsächlichen Verantwortung hinsichtlich des Lernprozesses. Während der Instruktivismus dem Lehrenden die Verantwortung für die Planung des Unterrichts zuschreibt, was die hohe Anforderung an ihn stellt, für jeden Schüler immer die optimalen Lernschritte kennen zu müssen, bezieht der Konstruktivismus den Lernenden mehr in das eigene Lernen ein. Der Lernprozess soll sich an den Interessen der Schüler orientieren und autonomes beziehungsweise entdeckendes Lernen fördern.⁶⁸⁶

Bernd Rüschoff schreibt in Bezug auf das konstruktivistische Lernen:

„Konstruktivistisches Lernen setzt einen reichen Lernkontext voraus, und Lernen sollte somit immer aus dem Kontext und dessen Lernrelevanz und Echtheit motiviert werden. Dabei spielen multimodales und multiperspektivisches Lernen eine besondere Rolle. All dies sind Konzeptionen,

⁶⁸³ Vgl. Grünewald/ Küster (2009): op. cit., S. 122.

⁶⁸⁴ Vgl. Fäcke (2011): op. cit., S. 54.

⁶⁸⁵ Vgl. Grünewald/ Küster (2009): op. cit., S. 122-123.

⁶⁸⁶ Vgl. Schlickau, Stephan (2009): *Neue Medien in der Sprach- und Kulturvermittlung*. Frankfurt am Main: Lang, S. 53-55.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

deren Umsetzung über eine durch technologiegestützte Werkzeuge bereicherte Lernumgebung möglich erscheint [...].“⁶⁸⁷

Er fordert daher einen forschungsorientierten Ansatz für das Fremdsprachenlernen und folgert:

„Für das Lernen und Erwerben einer Fremdsprache bedeutet dies, daß [sic!] Sprachlernen nicht einfach auf die Notwendigkeit des Umgangs mit der Zielsprache in kommunikativen Situationen beschränkt werden darf, sondern daß [sic!] das eigenständige Auseinandersetzen durch die Lernenden mit Sprache für den Erwerb von Sprachkompetenz ebenso wichtig ist. In diesem Zusammenhang kommt der Rolle der Lernenden nicht nur als aktiv an Kommunikation Beteiligte, sondern vor allem als selbständig handelnde und eigenverantwortlich entdeckende und experimentierende ‚Forscher‘ große Bedeutung zu (vgl. Rüschhoff & Wolff 1991). [...] Und ein Mehr an Aufgabenstellungen und Lerneraktivitäten im Sinne eines *discovery based learning* oder besser einer *research driven acquisition* erscheint für das Sprachlernen notwendig. Nur so können Sprach- und Lernbewußtheit [sic!] gefördert werden.“⁶⁸⁸

Nach Rüschoffs Ansicht ist also gerade der selbstständige und forschende Umgang mit der Sprache in aktiven Lernphasen zu fördern. Eben diese aktive Beschäftigung bietet die Auseinandersetzung mit realen Texten anhand von Korpora.

Beim konstruktivistischen Lernprozess stehen also nicht mehr gestellte traditionelle Grammatikübungen und Vokabeldrills im Vordergrund, sondern die selbstständige Suche nach Vokabular, Wortverbindungen und Wortfamilien, sowie unter Umständen das Forschen in Korpora, um anhand realer Materialien vorher aufgestellte grammatische Hypothesen zu bestätigen oder zu verwerfen. Die Schüler sollen selbst Zusammenhänge erschließen und sich durch ihre eigenständige Tätigkeit Wissen aneignen, welches sie dann mit ihrem bisherigen Wissen verknüpfen und darin verankern können.

Auch Rautenhaus beschreibt den Fremdsprachenerwerbsprozess als konstruktivistischen Verstehensprozess. Die Schüler lernen durch Erstellung und Verstärkung von Assoziationen, also durch die Einbettung neuer Information in das bereits vorhandene Wissensgefüge. Diesen Prozess müssen alle Schüler selbst durchführen, er kann ihnen von niemandem abgenommen werden. Sie müssen sich die Bereitschaft, neue Informationen aufzunehmen und in den bereits vorhandenen

⁶⁸⁷ Rüschhoff, Bernd (1997): „Multimedia Autorentools und Authentizität im Fremdsprachenlernen“. In: Siebold, J. (Hrsg.) (1997): *Sprache und Medien im Fremdsprachenunterricht. Beiträge des 3. Mediendidaktischen Kolloquiums Oktober 1996 in Rostock*. Rostock: Universitätsdruckerei Rostock, S. 66.

⁶⁸⁸ Rüschhoff (1997): op. cit., S. 67.

Wissensbeständen zu verankern, erst erarbeiten. Daher versteht Rautenhaus Lehren als „Hilfe zur Gestaltung schülereigener Lernprozesse“.⁶⁸⁹

Weiter bekundet Rautenhaus, dass der Unterricht mit der althergebrachten Grammatikvermittlung, wie er bislang in den Schulen praktiziert wurde, nur wenig mit dieser Vorstellung des schülereigenen Lernprozesses zu tun hat. Sie lehnt die deduktive Methode des Einordnens sprachlicher Phänomene in ein vorgegebenes Raster ab, da die Schüler dieses Raster selbst erstellen und eine individuelle, für sie verständliche Grammatik erarbeiten sollten. Genau hier sieht sie die Vorteile des *data-driven learning*, welches Gelegenheit zu autonomem, forschendem Lernen gibt. Hier steht der Lerner dem Sprachmaterial direkt, ohne die Vermittlung durch den Lehrer, das Lehrbuch, die Grammatik oder das Lexikon gegenüber und leitet die Regelmäßigkeiten der Sprache selbstständig und vor allem induktiv ab.⁶⁹⁰

Auch Schlickau gibt an, dass den Schülern in den aktuellen Sprachlerntheorien eine aktivere Rolle zukommt und es die Aufgabe der Lehrkraft ist, eine förderliche Atmosphäre zu schaffen sowie die Lernenden zu selbstständigem Lernen zu befähigen. In den westlichen Gesellschaften ist man sich einig darüber, dass die Befähigung zum selbstständigen Lernen auch im Hinblick auf die Sprach- und Kulturvermittlung ein wichtiges Ziel ist. Laut Schlickau scheint es jedoch nicht sinnvoll, ausnahmslos jeden Lernstoff selbstentdeckend erarbeiten zu lassen. Dies sei schon allein eine Frage der Zeit sowie des Vorwissens und der Dispositionen der Schüler. So sind für ihn durchaus bei manchem Lernstoff *pattern practice* oder das Lehren kognitiver Regeln empfehlenswert. Er rät daher zu einem Methodenmix, wobei die jeweilige Methode in Abhängigkeit vom grammatischen Lernstoff und relevanten Lernerdispositionen gewählt werden sollte. Wie bereits erwähnt, soll die Arbeit mit Korpora lediglich als Ergänzung zur traditionellen Lehrbucharbeit dienen und abwechslungsreiche Phasen des selbstentdeckenden Lernens anhand realer Sprachdaten in den Unterricht bringen. Für eine solche Vorgehensweise müssen die Schüler allerdings bereits selbstständiges Lernen erlernt haben, das heißt, sie müssen wissen, wie Wissenslücken festgestellt und infolgedessen geschlossen werden

⁶⁸⁹ Rautenhaus, Heike (1997): „Authentische Texte und Konkordanzprogramme im Englischunterricht“. In: Siebold, J. (Hrsg.) (1997): *Sprache und Medien im Fremdsprachenunterricht. Beiträge des 3. Mediendidaktischen Kolloquiums Oktober 1996 in Rostock*. Rostock: Universitätsdruckerei Rostock, S. 142.

⁶⁹⁰ Vgl. Rautenhaus (1997): op. cit., S. 143.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

können.⁶⁹¹ Es ist jedoch fraglich, inwiefern Jugendliche dazu in der Lage sind, eigene Wissensdefizite zu erkennen und ihren Lernprozess in diese Richtung zu lenken.

Gilmore erstellte ein Schema zur Veranschaulichung der Auswirkungen von *Herausforderung* und *Unterstützung* auf den Lernprozess. Beim Entwurf von Aufgaben mit realem Material sollte die Lehrkraft sicherstellen, dass die Lerner nicht überfordert werden, so zum Beispiel, wenn sie Input nach Form und Bedeutung gleichzeitig analysieren sollen. Für gewöhnlich wird zunächst die Bedeutung untersucht und erst danach werden die formalen Aspekte analysiert. Entlang der zwei Dimensionen *Herausforderung* und *Unterstützung* können alle pädagogischen Aktivitäten beschrieben werden und ihre verschiedenen Kombinationen haben unterschiedliche Effekte auf den Lernprozess.⁶⁹²

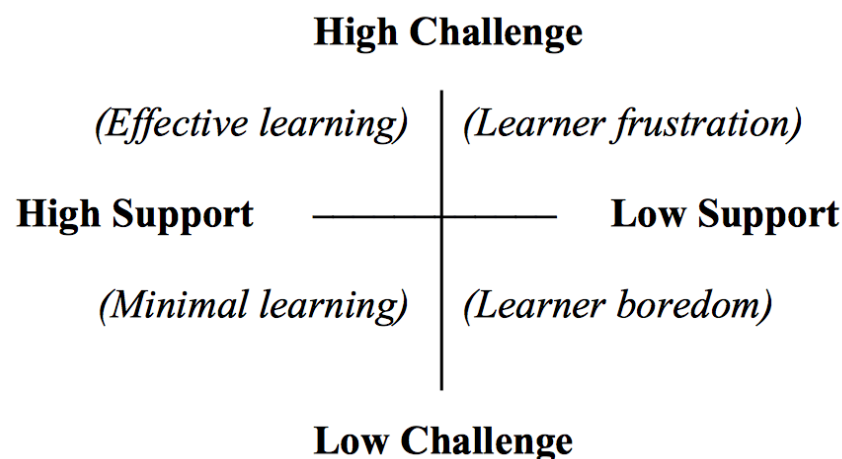


Abbildung 75: Schema zur Veranschaulichung der Auswirkungen von *Herausforderung* und *Unterstützung* auf den Lernprozess nach Gilmore (2007).⁶⁹³

Am effektivsten können Schüler demnach lernen, wenn sie viel gefordert, aber auch viel unterstützt werden. Dies ist der Fall, wenn die Herausforderung der Aufgabe gerade knapp über dem Können der Lerner liegt, sodass sie nur mit Unterstützung gelöst werden kann.⁶⁹⁴ Bei der Arbeit mit Korpora sollte die Lehrkraft die Schüler ebenso in einem ausgewogenen Verhältnis fördern und fordern, damit diese motiviert und weder über- noch unterfordert sind. Vor allem in Bezug auf die Einführung in die Korpusarbeit und die Erklärung der Funktionsweise der Korpora sind die Lerner auf den Lehrer

⁶⁹¹ Vgl. Schlickau (2009): op. cit., S. 52-53, 57-61.

⁶⁹² Vgl. Gilmore, Alex (2007): „Authentic materials and authenticity in foreign language teaching“. In: *Language Teaching* 40/2, S. 111-112.

⁶⁹³ Gilmore (2007): op. cit., S. 112.

⁶⁹⁴ Vgl. Gilmore (2007): op. cit., S. 112.

angewiesen. Dieser sollte ihnen jedoch viel Spielraum lassen und ihnen die Möglichkeit geben, selbst im Korpus Suchanfragen durchzuführen.

Auch Conolly ist der Meinung, die Ansicht, dass den Schülern immer nur Stückchen der Sprache vermittelt werden sollten, sodass sie nur mit diesen sprachlichen Strukturen konfrontiert werden, da sie nicht mit zu viel Neuem überhäuft werden sollten, ist nicht mehr gültig. Schüler können durchaus neue Strukturen passiv wahrnehmen und interpretieren, bevor sie lernen, diese aktiv zu gebrauchen. Ansonsten werde der Lernfortschritt der Schüler unnötig verzögert, da diese ihre rezeptiven Fähigkeiten genauso wie ihre produktiven Fähigkeiten trainieren müssen. Conolly räumt allerdings ein, dass aufgrund der Quantität der neuen sprachlichen Informationen vieles, was die Lerner wahrnehmen, vage und teilweise unbewusst bleibt und im Unterricht auch nicht bewusst gemacht wird.⁶⁹⁵ Demnach ist die Tatsache, dass die Schüler bei der Arbeit mit Korpora zwangsläufig auf unbekanntes Vokabular und noch nicht behandelte sprachliche Strukturen stoßen, nicht durchweg negativ zu sehen, sondern sie ist unter Umständen sogar förderlich für die Entwicklung ihrer rezeptiven Fähigkeiten.

Laut Siepmann können Konkordanzprogramme einen wichtigen Beitrag zur Förderung des sogenannten *noticing* leisten, worunter man die bewusste Wahrnehmung versteht.⁶⁹⁶ Je stärker Lerner das sprachliche Material, das ihnen durch den Lehrer, durch Medien oder durch ein Gespräch mit Muttersprachlern vermittelt wird, bewusst wahrnehmen, desto eher wird dieser fremdsprachliche *Input* zu einem Teil des *Intake*, also der Lernaltersprache.⁶⁹⁷ Da die Konkordanzanalyse ein induktiv-entdeckendes Lernen darstellt, führt sie zu einer tieferen kognitiven Verarbeitung des sprachlichen Materials.⁶⁹⁸

Wie an obigen Ausführungen zu sehen ist, kann anhand der Arbeit mit Korpora einigen Forderungen der Ansätze und Methoden des modernen Fremdsprachenunterrichts nachgekommen werden. Im Folgenden soll nun überprüft werden, ob die Korpusarbeit auch mit dem *Gemeinsamen Europäischen Referenzrahmen für Sprachen* kompatibel ist.

⁶⁹⁵ Vgl. Conolly, Declan B. (1996): „Are there degrees of authenticity?“. In: *Zielsprache Englisch*, 1996, 26 (2), S. 15-16.

⁶⁹⁶ Vgl. Siepmann (2009): op. cit., S. 326.

⁶⁹⁷ Vgl. Schmidt, Richard (1994): „Deconstructing consciousness in search of useful definitions for applied linguistics“. In: *AILA Review*, 11, S. 17-18.

⁶⁹⁸ Vgl. Siepmann (2009): op. cit., S. 326.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

3.1.2. Der Gemeinsame Europäische Referenzrahmen für Sprachen

Neben dem jeweiligen Lehrplan hat sich der Sprachunterricht seit dem Jahr 2001 auch nach dem *Gemeinsamen Europäischen Referenzrahmen für Sprachen* (GER) zu richten. Dieser überstaatliche Impuls wurde vom Europarat mit der Zielsetzung verabschiedet, den Sprachunterricht hinsichtlich Lehrplänen, Prüfungen und Lehrmaterialien zu homogenisieren und transparenter zu gestalten.⁶⁹⁹ Der GER ist ein Instrument, das Transparenz und Vergleichbarkeit in Bezug auf die Ziele, die Gestaltung und die Ergebnisse des Lernprozesses schaffen soll, um Hindernisse zwischen verschiedenen Bildungssystemen überwinden und die Zusammenarbeit der Bildungseinrichtungen erleichtern zu können.⁷⁰⁰

Im Referenzrahmen werden diejenigen Kenntnisse und Fähigkeiten beschrieben, die zur erfolgreichen Kommunikation in der jeweiligen Sprache benötigt werden, wobei auch kulturelle Kontexte dargestellt sind. Dabei gibt es sechs Kompetenzniveaus, welche auf die bereits erworbenen Fähigkeiten ausgerichtet sind und für Lernende somit eine motivierende Rückmeldung darstellen.⁷⁰¹ Die Niveaustufen des GER umfassen drei grundlegende Levels (A-C), welche sich nochmals in insgesamt 6 Stufen des Sprachniveaus unterteilen (A1-C2).⁷⁰² Im Niveau A1 ‚Anfänger‘ kann der Lerner folgende Kenntnisse und Fähigkeiten nachweisen:

„Kann vertraute, alltägliche Ausdrücke und ganz einfache Sätze verstehen und verwenden, die auf die Befriedigung konkreter Bedürfnisse zielen. Kann sich und andere vorstellen und anderen Leuten Fragen zu ihrer Person stellen – z. B. wo sie wohnen, was für Leute sie kennen oder was für Dinge sie haben – und kann auf Fragen dieser Art Antwort geben. Kann sich auf einfache Art verständigen, wenn die Gesprächspartnerinnen oder Gesprächspartner langsam und deutlich sprechen und bereit sind zu helfen.“⁷⁰³

Der Referenzrahmen orientiert sich für die Beschreibung also sowohl an der Komplexität der sprachlichen Äußerungen als auch an den Kommunikationsabsichten und Themen, welche der Lerner auf dieser Stufe versprachlichen kann. Auf dem Niveau A1 sind die Themen beispielsweise ‚Ich‘, ‚Meine Familie und Freunde‘, ‚Mein Haus‘ und ‚Mein Zimmer‘. Neben der Beschreibung als Kompetenzen, welche eine positive

⁶⁹⁹ Vgl. Fäcke (2011): op. cit., S. 72-73.

⁷⁰⁰ Vgl. Grünwald/ Küster (2009): op. cit., S. 134.

⁷⁰¹ Vgl. Fäcke (2011): op. cit., S. 72-73.

⁷⁰² Vgl. Europarat (ed.2001): Gemeinsamer Europäischer Referenzrahmen für Sprachen (GER). Online Publikation: <http://www.europaecischer-referenzrahmen.de> (06.09.2017).

⁷⁰³ Europarat (ed.2001): op. cit. (06.09.2017).

Wertschätzung vermitteln, wird hier nicht konkret angegeben, welche sprachlichen Strukturen verwendet werden sollen, sondern welche sprachlichen Absichten und Kommunikationssituationen die Schüler mit ihrem Wissen meistern können.

Diese Beschreibungen lassen zwar viel Spielraum, sind aber ohne Unterrichtserfahrung schwer einzuschätzen, weshalb die Vereinbarkeit mit der Korpusarbeit zunächst schwierig, um nicht zu sagen unmöglich scheint, da die Korpora mit den darin enthaltenen Texten zu komplex für den Anfangsunterricht scheinen. Es gibt jedoch nicht nur Referenzkorpora oder Korpora der Zeitungssprache, wie bisher zu sehen war, sondern auch pädagogische Korpora. Das multimodale pädagogische Korpus *Sacodeyl*⁷⁰⁴ (vgl. Kapitel 2.4.3.1.) lässt beispielsweise nicht nur Suchen nach sprachlichen Strukturen zu, sondern auch solche nach Themenbereichen und den Niveaustufen des GER. Im *Sacodeyl* finden sich Themen wie *Presentarse y hablar sobre uno mismo* oder *Mi casa*, welche somit beide laut der Beschreibung des GER für Lerner mit der Kompetenz A1 genutzt werden können. Zudem ermöglicht das *Sacodeyl* die explizite Suche nach Texten für eine bestimmte Niveaustufe, wodurch die Lehrer speziell nach Interviews für ihre jeweilige Lerngruppe suchen können. Es umfasst Interviews in den Niveaustufen A1 bis B2.⁷⁰⁵ Im multimodalen Korpus *Backbone Spanish* (vgl. Kapitel 2.4.3.2.) werden die dort aufgelisteten Interviews ebenso nach den Niveaustufen des GER eingeteilt, wodurch sich dieses Korpus auch für die Verwendung im Unterricht auszeichnet.

Im bayerischen Lehrplan für das G8 wird in der 8. Jahrgangsstufe, also dem ersten Lernjahr Spanisch als dritte Fremdsprache, die Stufe A2 des GER erreicht.⁷⁰⁶ Selbiges trifft für die 10. Jahrgangsstufe im Fach Spanisch als spätbeginnende Fremdsprache zu.⁷⁰⁷ Somit können im ersten Lernjahr bereits einige Texte des Korpus verwendet werden.

Für das Niveau B1 ‚Fortgeschrittene Sprachverwendung‘ wird folgende Beschreibung im GER angegeben:

⁷⁰⁴ Vgl. Pérez-Paredes: *Sacodeyl. European Youth Language*. Op. cit. (18.11.2016).

⁷⁰⁵ Vgl. Pérez-Paredes, Pascual: *Sacodeyl. Category Tree*. Online Publikation: <http://sacodeyl.inf.um.es/sacodeyl-search2/faces/search.jsp> (06.09.2017).

⁷⁰⁶ Vgl. Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Spanisch (3. Fremdsprache) – Jahrgangsstufe 8*. Online Publikation: <http://www.isb-gym8-lehrplan.de/contentserv/3.1.neu/g8.de/index.php?StoryID=26253> (06.09.2017).

⁷⁰⁷ Vgl. Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Spanisch (spätbeginnend) – Jahrgangsstufe 10*. Online Publikation: <http://www.isb-gym8-lehrplan.de/contentserv/3.1.neu/g8.de/index.php?StoryID=26804> (06.09.2017).

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

„Kann die Hauptpunkte verstehen, wenn klare Standardsprache verwendet wird und wenn es um vertraute Dinge aus Arbeit, Schule, Freizeit usw. geht. Kann die meisten Situationen bewältigen, denen man auf Reisen im Sprachgebiet begegnet. Kann sich einfach und zusammenhängend über vertraute Themen und persönliche Interessengebiete äußern. Kann über Erfahrungen und Ereignisse berichten, Träume, Hoffnungen und Ziele beschreiben und zu Plänen und Ansichten kurze Begründungen oder Erklärungen geben.“⁷⁰⁸

Neben dem verbesserten Verständnis und den weiterentwickelten Fähigkeiten in schriftlicher und mündlicher Sprachproduktion hat der Lerner nun das Wissen erworben, sich über Themen wie Arbeit, Schule, Freizeit, Ereignisse, Träume, Hoffnungen und Ziele auszutauschen sowie sich auf Reisen zu verständigen.

Auch zu diesen Themen finden sich Texte im Korpus *Sacodeyl*, wie die Auswahlkriterien *Hobbies*, *Instituto*, *Planes futuros* oder *Viajar* zeigen.⁷⁰⁹ Dieses Kompetenzniveau wird am bayerischen Gymnasium im Fach Spanisch als dritte Fremdsprache allgemein in der 10. Klasse, also im dritten Lernjahr, erreicht. Im Bereich Leseverstehen können sich Schüler in dieser Jahrgangsstufe sogar auf das Niveau B1+ steigern.⁷¹⁰ Bei Spanisch als spätbeginnender Fremdsprache wird das Niveau B1 im Hinblick auf das Leseverstehen in der 11. Jahrgangsstufe erreicht und im Bereich der kommunikativen Fähigkeiten und sprachlichen Mittel in der 12. Jahrgangsstufe.⁷¹¹

Im Fach Spanisch als dritte Fremdsprache wird am Ende der 12. Jahrgangsstufe im Bereich der kommunikativen Fertigkeiten das Kompetenzniveau B2 und im Hinblick auf das Leseverstehen sogar die Stufe C1 erlangt.⁷¹² Die Beschreibung des Niveaus C1 ‚Fachkundige Sprachkenntnisse‘ lautet:

„Kann ein breites Spektrum anspruchsvoller, längerer Texte verstehen und auch implizite Bedeutungen erfassen. Kann sich spontan und fließend ausdrücken, ohne öfter deutlich erkennbar nach Worten suchen zu müssen. Kann die Sprache im gesellschaftlichen und beruflichen Leben oder

⁷⁰⁸ Europarat (ed.2001): op. cit. (06.09.2017).

⁷⁰⁹ Vgl. Pérez-Paredes: *Sacodeyl. European Youth Language*. Op. cit. (18.11.2016).

⁷¹⁰ Vgl. Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Spanisch (3. Fremdsprache) – Jahrgangsstufe 10*. Online Publikation: <http://www.isb-gym8-lehrplan.de/contentserv/3.1.neu/g8.de/index.php?StoryID=26431> (06.09.2017).

⁷¹¹ Vgl. Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Spanisch (spätbeginnend) – Jahrgangsstufe 11*. Online Publikation: <http://www.isb-gym8-lehrplan.de/contentserv/3.1.neu/g8.de/index.php?StoryID=26805> (06.09.2017). Und: *Spanisch (spätbeginnend) – Jahrgangsstufe 12*. Online Publikation: <http://www.isb-gym8-lehrplan.de/contentserv/3.1.neu/g8.de/index.php?StoryID=26806> (06.09.2017).

⁷¹² Vgl. Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Spanisch (3. Fremdsprache) – Jahrgangsstufe 11/12*. Online Publikation: <http://www.isb-gym8-lehrplan.de/contentserv/3.1.neu/g8.de/index.php?StoryID=26514> (06.09.2017).

in Ausbildung und Studium wirksam und flexibel gebrauchen. Kann sich klar, strukturiert und ausführlich zu komplexen Sachverhalten äußern und dabei verschiedene Mittel zur Textverknüpfung angemessen verwenden.“⁷¹³

Ein Lerner mit dieser Kompetenz kann also anspruchsvolle, längere Texte verstehen und somit durchaus mit verschiedenen Korpora arbeiten. Den Schülern der Qualifikationsstufe ist es damit möglich, die verschiedenen Korpora der spanischen Sprache in vollem Umfang zu nutzen, sei es zur Überprüfung der Verwendung bestimmter sprachlicher Strukturen oder zur Suche von Texten zu einem bestimmten Themenkomplex.

Zudem betont der GER das autonome Lernen, indem viel Wert auf den Aufbau lernstrategischen Wissens gelegt wird, was dazu führt, dass im Unterricht Lerntechniken thematisiert sowie reale Materialien verwendet werden. Außerdem erkennt er den Motivationsgehalt eben dieses realen Lernmaterials für die Schüler an. Grundsätzlich geht es im GER darum, dass die Lerner so früh wie möglich in der Fremdsprache kommunizieren, um bereits in einem frühen Lernstadium Erfolgserlebnisse zu erzielen.⁷¹⁴

Demnach können Korpora durchaus in Übereinstimmung mit dem GER und zur Förderung der Erreichung der Niveaustufen sowie einiger Forderungen des GER im Spanischunterricht an bayerischen Gymnasien genutzt werden. Wie bereits mehrfach betont, handelt es sich bei Korpora um Sammlungen realer Daten und die Korpusarbeit fördert durch die aktive Beteiligung des Lerners am eigenen Lernprozess zudem die Lernerautonomie.

3.1.3. Der Lehrplan für Spanisch an bayerischen Gymnasien

Nach der Besprechung der Vereinbarkeit der Korpusarbeit mit dem *Gemeinsamen Europäischen Referenzrahmen für Sprachen*, welcher übergeordnet fungiert, wird im Folgenden der bayerische Lehrplan für die moderne Fremdsprache Spanisch auf dessen Kompatibilität mit der Korpusarbeit überprüft. Dieser Vergleich soll als ein Beispiel stellvertretend für andere Bundesländer und Schularten durchgeführt werden.

Die vorliegende Arbeit orientiert sich am Lehrplan für die achtjährige Gymnasiallaufbahn (G8), da zum Zeitpunkt des Verfassens der Arbeit der Lehrplan Plus gerade in der 5. Jahrgangsstufe eingeführt und die erneute Erweiterung der gymnasialen

⁷¹³ Europarat (ed.2001): op. cit. (06.09.2017).

⁷¹⁴ Vgl. Grünewald/ Küster (2009): op. cit., S. 134.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

Laufbahn auf neun Jahre soeben erst beschlossen wurde. Derzeit wird die Fremdsprache Spanisch am bayerischen Gymnasium ab der 8. Klasse (Spanisch als 3. Fremdsprache) beziehungsweise ab der 10. Klasse (Spanisch als spätbeginnende Fremdsprache) unterrichtet, weshalb neue Lehrpläne diese Klassenstufen erst in frühestens drei Jahren betreffen würden.

Im *Fachprofil Moderne Fremdsprachen* für das bayerische Gymnasium steht unter der Überschrift ‚Besonderheiten der modernen Fremdsprachen‘ Folgendes geschrieben:

„Im Unterricht sollen die Schüler grundsätzlich die Zielsprache verwenden. Zum Einhalten einer lexikalischen und grammatischen Progression ist in der Grundphase eine Orientierung des Unterrichts an den zugelassenen Lehrwerken hilfreich; das Lehrbuch ist jedoch nicht dem Lehrplan gleichzusetzen. Zur Förderung von interkulturellem Lernen und situationsgerechtem Sprachhandeln begegnen die Schüler vom ersten Lernjahr an – ggf. auch ergänzend zum Lehrbuch – motivierenden authentischen Materialien. Übungsformen und Leistungserhebungen sollen so vielfältig sein, dass sie im Lauf eines Schuljahres die vom Lehrplan geforderten unterschiedlichen Kenntnisse und Fertigkeiten sowie das Grundwissen einbeziehen.“⁷¹⁵

Die Schüler sollen also vom ersten Lernjahr an „motivierenden authentischen Materialien“ begegnen. Es ist somit durchaus im Sinne des Lehrplans, wenn im Unterricht mit Korpora gearbeitet wird, die aus realen Texten bestehen. Zudem ist eine Orientierung an den zugelassenen Lehrwerken hilfreich, das Lehrbuch ist aber nicht dem Lehrplan gleichzusetzen. Bisher sind in den Lehrwerken allerdings hauptsächlich didaktisierte, also von Lehrbuchautoren verfasste Texte zu finden. Die Lehrbucharbeit allein kommt daher dieser Forderung des Lehrplans nicht nach.

Außerdem sollen die Übungsformen vielfältig sein und unterschiedliche Kenntnisse und Fertigkeiten sowie das Grundwissen miteinbeziehen. Übungen, die auf realen Korpusdaten aufbauen, können sowohl die eben gelernten Vokabeln sowie die neue Grammatik einüben, als auch das Grundwissen prüfen. Um beispielsweise Konkordanzen aus Korpora zu untersuchen und daraus Regelmäßigkeiten abzuleiten, müssen die Schüler sowohl auf ihr Grundwissen bauen, als auch auf ihre Fertigkeit, Sätze und Texte trotz Wortschatzlücken zu verstehen und zu interpretieren. Diese Art des Lernens anhand realer Textausschnitte, bei welchen den Schülern auch ein Teil des Wortschatzes fehlt, stellt eine realistische Sprachsituation dar.

⁷¹⁵ Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Fachprofil Moderne Fremdsprachen*. Online Publikation: <http://www.isb-gym8-lehrplan.de/contentserv/3.1.neu/g8.de/index.php?StoryID=26366> (08.09.2017).

Des Weiteren heißt es im Punkt *Unterrichtsqualität* in ‚Das Gymnasium in Bayern‘, dass die Lehrkräfte „die Lernsituationen methodisch vielfältig“ gestalten sowie den Schülern helfen sollen, „sich Wissen zunehmend eigenständig zu erschließen und damit verbundene Fertigkeiten nachhaltig anzueignen“. ⁷¹⁶ Durch die Arbeit mit Korpora wird der Unterricht in jedem Fall abwechslungsreicher, wobei gleichzeitig die Selbstständigkeit der Lerner aufgrund der aktiven Beteiligung am Lernprozess gefördert wird.

Im *Fachprofil Spanisch* des Staatsinstituts für Schulqualität und Bildungsforschung München wird angegeben, dass die Schüler Einblicke in Geographie, Gesellschaft, Politik, Wirtschaft und andere Bereiche der Kulturen Spaniens und Hispanoamerikas gewinnen sollen. ⁷¹⁷ Derartige Einblicke werden bestmöglich über reale Texte, beispielsweise aus Sprachkorpora, vermittelt. Auch die Besonderheiten der jeweiligen Kultur und Unterschiede im Vergleich zum deutschsprachigen Kulturraum können anhand des Sprachmaterials festgestellt werden.

Zu den jeweiligen regionalen Varianten schreibt das Staatsinstitut für Schulqualität und Bildungsforschung München:

„Die von den Schülern produktiv zu erlernende sprachliche Form soll dem peninsularen Spanisch oder aber einem der hispanoamerikanischen Standards entsprechen. Im rezeptiven Bereich begegnen die Schüler Varianten aus dem gesamten spanischen Sprachraum.“ ⁷¹⁸

Auch die Erfüllung dieser Vorgabe wird durch die Korpusarbeit erleichtert. Wie im zweiten Kapitel zu sehen war, beinhalten die großen Referenzkorpora (siehe Kapitel 2.1.1.) Texte des peninsularen sowie lateinamerikanischen Spanisch, welche somit zum Erlernen des Standards genutzt werden können. Sie beinhalten meist auch Texte der gesprochenen Sprache. Zudem stehen eigens Korpora der gesprochenen Sprache zur Verfügung, anhand derer Unterschiede zwischen gesprochenem und geschriebenem Spanisch veranschaulicht werden können. Des Weiteren gibt es Parallelkorpora wie das CLUVI (siehe Kapitel 2.3.2.), anhand dessen die Varietäten des Galicischen, Baskischen

⁷¹⁶ Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Das Gymnasium in Bayern*. Online Publikation: <http://www.isb-gym8-lehrplan.de/contentserv/3.1.neu/g8.de/index.php?StoryID=26350> (07.09.2017).

⁷¹⁷ Vgl. Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Fachprofil Spanisch*. Online Publikation: <http://www.isb-gym8-lehrplan.de/contentserv/3.1.neu/g8.de/index.php?StoryID=26376> (08.09.2017).

⁷¹⁸ Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Fachprofil Spanisch*. Op. cit. (08.09.2017).

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

und Katalanischen mit dem Kastilischen, aber auch mit dem Englischen und Französischen verglichen werden können. Die Veranschaulichung der Existenz sprachlicher Varianten wird somit ebenfalls erleichtert.

In Bezug auf die Verwendung der Fremdsprache zu Beginn des Fremdspracherwerbs findet man im Lehrplan für Spanisch als dritte Fremdsprache in der Jahrgangsstufe 8 folgende Informationen:

„Der Unterricht in Spanisch als dritter Fremdsprache baut auf Kenntnissen, Fertigkeiten und Grundhaltungen auf, welche die Schüler beim Erlernen der ersten und zweiten Fremdsprache erworben haben. Grundsätzlich findet der Unterricht in der Zielsprache statt, wobei die Vorkenntnisse einen raschen Fortschritt ermöglichen, der mit motivierenden Erfolgserlebnissen verbunden ist.“⁷¹⁹

Während die Arbeit mit spanischsprachigen Korpora oder daraus stammenden Materialien wohl im Unterricht der ersten oder zweiten Fremdsprache im ersten Lernjahr nicht möglich wäre, baut der Unterricht in der 3. Fremdsprache bereits auf einem gewissen Repertoire an Wissen, Fähigkeiten und Fertigkeiten auf, was eine schnellere Progression ermöglicht. Grundsätzlich erleichtert das bis zur Jahrgangsstufe 8 beziehungsweise 10 erworbene Wissen über Fremdsprachen wie Englisch, Französisch oder Latein das Erlernen der dritten Fremdsprache und ermöglicht damit den kommunikativen Austausch in der Fremdsprache von Beginn an. Konzepte wie Wortarten, Synonymie, Wortfamilien oder andere grammatische Kategorien sind bereits hinreichend bekannt oder können anhand des bis dahin Gelernten leicht erschlossen werden. Daher kann bereits im ersten Lernjahr, wenn auch gegen Ende desselben, mit Korpora gearbeitet werden, wie im vorherigen Kapitel zum *Gemeinsamen Europäischen Referenzrahmen für Sprachen* (siehe Kapitel 3.1.2.) zu sehen war. Das Niveau A2 wird bereits im ersten Lernjahr erreicht und somit können beispielsweise die Korpora *Sacodeyl* oder *Backbone Spanish* genutzt werden.

Der bayerische Lehrplan für Gymnasien ist für das Fach Spanisch in jeder Jahrgangsstufe in vier Bereiche unterteilt, wobei der erste den größten Umfang der Beschreibung einnimmt: Sprache, Umgang mit Texten und Medien, interkulturelles Lernen und Landeskunde sowie Lernstrategien und Methoden selbstständigen Arbeitens.

⁷¹⁹ Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Spanisch (3. Fremdsprache) – Jahrgangsstufe 8*. Op. cit. (06.09.2017).

Aus diesem Grund werden die folgenden Ausführungen in der gleichen Weise eingeteilt und betitelt.

3.1.3.1. Sprache

Unter dem Titel ‚Sprache‘ werden im bayerischen Lehrplan sowohl die kommunikativen Fertigkeiten Hörverstehen, mündliche Ausdrucksfähigkeit, Leseverstehen, schriftliche Ausdrucksfähigkeit und Sprachmittlung als auch sprachliche Mittel wie Aussprache, Intonation, Wortschatz, Idiomatik und Grammatik sowie die Sprachreflexion verstanden. Es handelt sich somit um Sprache im Sinne von Sprachfertigkeit.

In Bezug auf das Hörverstehen sollen die Schüler im ersten Lernjahr bereits „einfache, in gemäßigtem Tempo gesprochene und deutlich artikuliert Äußerungen in alltäglichen Kommunikationssituationen global und in wichtigen Details verstehen, auch in Gesprächen unter Muttersprachlern, ggf. mithilfe gelegentlicher Rückfragen“ und „auch regional leicht gefärbte Äußerungen von Muttersprachlern zu vertrauten Themen in Grundzügen verstehen“. ⁷²⁰ Diesbezüglich können Korpora gesprochener Sprache eine große Hilfe sein. Sie ermöglichen einen einfachen und problemlosen Zugang zu realem, muttersprachlichem Material und enthalten teilweise auch Texte mit verschiedenen regionalen Färbungen. Des Weiteren sollen die Schüler „die Intonation grundlegender Satzmuster erkennen“, wobei muttersprachliche Hörtexte eine realistischere Vorlage bieten als beispielsweise eigens für den Fremdsprachenunterricht aufgezeichnete Lehrbuchtexte. ⁷²¹ Bei Schülern im Anfangsunterricht, wie es in der 8. Jahrgangsstufe der Fall ist, oder auch bei fortgeschrittenen Lernern, sofern Interviews eine dialektale Färbung aufweisen, ist die Möglichkeit, den Text in den Korpora *Sacodeyl* oder *Backbone Spanish* parallel mitzulesen, sehr nützlich, allerdings kann die Lehrkraft den Schülern ebenso gut nur den Hörtext vorspielen, ohne sie mitlesen zu lassen, und ihnen Fragen dazu stellen. Die Schüler können sich die Texte auch selbst anhören und mitlesen und im Folgenden die muttersprachliche Intonation nachahmen. Die Angabe des jeweiligen Kompetenzniveaus in diesen beiden Korpora erleichtert dabei die Suche nach passenden

⁷²⁰ Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Spanisch (3. Fremdsprache) – Jahrgangsstufe 8*. Op. cit. (06.09.2017).

⁷²¹ Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Spanisch (3. Fremdsprache) – Jahrgangsstufe 8*. Op. cit. (06.09.2017).

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

Interviews. Die Themenbereiche im *Sacodeyl* sind unter anderem *Mi casa*, *Rutinas y Actividades cotidianas*, *Hobbies*, *Vacaciones* oder *Instituto* und decken sich somit mit denen des Lehrplans für das erste Lernjahr: „Alltag, Familie, Herkunft, Wohnung, Schule, Kleidung, Freizeit, Reise“.⁷²²

Im Fach Spanisch als spätbeginnende Fremdsprache sollen die Lerner in der 12. Klasse, also dem 3. Lernjahr, „längere und komplexere Gespräche und Äußerungen von Muttersprachlern in normalem Sprechtempo global und in wichtigen Details verstehen (auch z. B. regionale Varianten, Jugendsprache)“.⁷²³ Hierfür würde sich zum einen das Korpus *Sacodeyl* eignen, welches, wie bereits erwähnt, Interviews von Jugendlichen zwischen 13 und 18 Jahren zu Alltagsthemen enthält, und zum anderen das COLA, ein Korpus der Jugendsprache (siehe Kapitel 2.1.2.1.).

Hinsichtlich des Leseverstehens sollen die Schüler bereits im 1. Lernjahr „einfache und klar strukturierte, ggf. auch authentische Texte zu Themen aus dem Alltagsbereich und von landeskundlichem Interesse global und in wichtigen Details verstehen“ (vgl. 3.2.1. zur Diskussion des Begriffs ‚authentisch‘).⁷²⁴ Auch hier können die Korpora als Quellen realer Texte oder zumindest Textausschnitte dienen, sofern die dort behandelten Themen nicht zu komplex oder abstrakt sind.

In der 9. Jahrgangsstufe sollen die Schüler in Bezug auf das Leseverstehen „etwas komplexere und längere, zunehmend authentische Texte auch mit teilweise unbekanntem Vokabular zu vertrauten Themen global verstehen (kursorisch lesen)“ sowie „verschiedenartigen, auch längeren und etwas komplexeren authentischen Texten zu vertrauten Themen auf eine bestimmte Fragestellung hin wichtige Details entnehmen (selektiv lesen)“.⁷²⁵ Somit wird die zunehmende Verwendung realer Texte bereits ab dem zweiten Lernjahr gefordert.

Im Hinblick auf die sprachlichen Mittel sollen die Lerner in der Qualifikationsphase „Aussprache, Intonation und Sprechgeschwindigkeit einer authentischen Norm

⁷²² Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Spanisch (3. Fremdsprache) – Jahrgangsstufe 8*. Op. cit. (06.09.2017).

⁷²³ Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Spanisch (spätbeginnend) – Jahrgangsstufe 12*. Op. cit. (06.09.2017).

⁷²⁴ Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Spanisch (3. Fremdsprache) – Jahrgangsstufe 8*. Op. cit. (06.09.2017).

⁷²⁵ Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Spanisch (3. Fremdsprache) – Jahrgangsstufe 9*. Online Publikation: <http://www.isb-gym8-lehrplan.de/contentserv/3.1.neu/g8.de/index.php?StoryID=26220> (06.09.2017).

annähern“ sowie „beim gestaltenden Lesen und Vortragen Intonation, Artikulation und flüssige Darbietung üben“. ⁷²⁶ Dafür bedarf es allerdings muttersprachlicher Vorbilder, doch spanische oder lateinamerikanische Lehrer sind im deutschsprachigen Raum selten, weshalb gewöhnlich auf auditive und audiovisuelle Medien zurückgegriffen wird. An diesem Punkt kommen erneut die Korpora *Sacodeyl* und *Backbone Spanish* ins Spiel (siehe Kapitel 2.4.3.1. und 2.4.3.2.), denn hier ist es möglich, sich ein Interview anzuhören und gleichzeitig den Text mitzulesen. Die Schüler können auf diese Weise auch selbstständig ihre Aussprache und Intonation üben, indem sie sich die sprachlichen Ausschnitte anhören und nachsprechen.

Im Lehrplan der 8. Jahrgangsstufe für Spanisch als 3. Fremdsprache und der 10. Jahrgangsstufe für Spanisch als spätbeginnende Fremdsprache wird für die Grammatik angegeben, dass die Schüler *ir a + infinitivo* und *estar + gerundio* erlernen. Anhand einer Auflistung von Konkordanzen zu einem dieser Themen können die Schüler selbst Regelmäßigkeiten erkennen und im Folgenden Regeln ableiten. Sollte für die Schüler des ersten Lernjahrs, egal ob in der 8. oder in der 10. Klasse, der Umgang mit Korpora noch zu schwierig sein, weil die Suche nach bestimmten Grammatikeinheiten zu komplex ist oder sie sich aufgrund ihres begrenzten Wortschatzes in den unbekannten Lexemen der Konkordanzen verlieren würden (siehe Kapitel 1.4.3.), so kann die Lehrkraft auch eine Reihe von Sprachbeispielen mit den jeweiligen Konstruktionen auswählen und ihnen als Arbeitsblatt vorlegen. Da die Schüler im ersten Lernjahr über einen relativ geringen Wortschatz verfügen, sollte die Lehrkraft einfache Konkordanzen auswählen, deren Wortschatz größtenteils bekannt ist. Auch der Gebrauch von *ser*, *estar* und *hay*, der in dieser Jahrgangsstufe erlernt werden soll, kann anhand der Gegenüberstellung von Konkordanzen verdeutlicht oder eingeübt werden (siehe hierfür auch Kapitel 4.1.7. sowie Anhang 21). ⁷²⁷

Des Weiteren sollen die Schüler der 10. Jahrgangsstufe im Fach Spanisch als spätbeginnende Fremdsprache häufig gebrauchte Formen der Vergangenheitstempora *indefinido* und *imperfecto* sowie erste Aspektunterschiede beim Zeitengebrauch kennen

⁷²⁶ Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Spanisch (3. Fremdsprache) – Jahrgangsstufe 11/12*. Op. cit. (06.09.2017).

⁷²⁷ Vgl. Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Spanisch (3. Fremdsprache) – Jahrgangsstufe 8*. Op. cit. (06.09.2017). Und: *Spanisch (spätbeginnend) – Jahrgangsstufe 10*. Op. cit. (06.09.2017).

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

und die Zeiten verwenden können.⁷²⁸ Auch hierbei können Konkordanzen die Gebrauchsregeln verdeutlichen und Vergleiche ermöglichen.

In Bezug auf die Sprachreflexion heißt es bereits im ersten Lernjahr, die Schüler sollen „sich der Existenz regionaler Varianten des Spanischen in Spanien und Hispanoamerika bewusst sein“.⁷²⁹ Die regionale Vielfalt der spanischen Sprache kann in vielerlei Hinsicht durch Korpora veranschaulicht werden. Zum einen können Vergleiche zwischen Korpora des peninsularen und solchen des lateinamerikanischen Spanisch durchgeführt werden. Zum anderen können auch die Varietäten des Katalanischen, Galicischen und Baskischen dem Kastilischen mithilfe des CLUVI (siehe Kapitel 2.3.2.) gegenübergestellt werden. Im Lehrplan für Spanisch als spätbeginnende Fremdsprache wird das Kennenlernen regionaler Varianten in den Jahrgangsstufen 11 und 12 gefordert.⁷³⁰

3.1.3.2. Umgang mit Texten und Medien

Der Bereich ‚Umgang mit Texten und Medien‘ umfasst im bayerischen Lehrplan sowohl die Texterschließung als auch die Texterstellung. Im vierten und fünften Lernjahr Spanisch als dritte Fremdsprache heißt es in diesem Bereich:

„Die Schüler vertiefen und erweitern ihre bisher erworbenen Erfahrungen und Techniken im Umgang mit Texten und wenden diese zunehmend selbständig an. [...] Geeignete Techniken und Strategien des Hör-, Hör-/Seh- und Leseverstehens, Methoden der Textanalyse und die Versprachlichung eigener Meinungen und Gedanken werden an einem breiten Spektrum von Themen und einem variablen Korpus authentischer Texte geübt und stetig verbessert. Die Fähigkeit zu differenziertem Textverständnis wird als wesentliche Voraussetzung für das Verfassen eigener, klar strukturierter, sach- und adressatengerechter Texte ausgebaut.“⁷³¹

Im Lehrplan für die Qualifikationsphase ist damit das selbstständige Arbeiten zunehmend von Bedeutung sowie der Umgang mit vielfältigen realen Texten, welche zum einen als Grundlage für die Entwicklung des Leseverstehens dienen, zum anderen

⁷²⁸ Vgl. Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Spanisch (spätbeginnend) – Jahrgangsstufe 10*. Op. cit. (06.09.2017).

⁷²⁹ Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Spanisch (3. Fremdsprache) – Jahrgangsstufe 8*. Op. cit. (06.09.2017).

⁷³⁰ Vgl. Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Spanisch (spätbeginnend) – Jahrgangsstufe 11*. Op. cit. (06.09.2017). Und: *Spanisch (spätbeginnend) – Jahrgangsstufe 12*. Op. cit. (06.09.2017).

⁷³¹ Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Spanisch (3. Fremdsprache) – Jahrgangsstufe 11/12*. Op. cit. (06.09.2017).

aber auch eine Voraussetzung für das Verfassen eigener Texte darstellen. Weiter heißt es:

„Durch den Umgang mit einem breiten Spektrum an authentischen, auch elektronischen Medien erlangen die Schüler zunehmende Sicherheit darin, Darstellungen aus unterschiedlichen Quellen im Hinblick auf ihre Verlässlichkeit sowie die Art der Informationsaufbereitung zu vergleichen, zu beurteilen und in ihrem Wert einzuschätzen. Einfachere wissenschaftliche Texte aus unterschiedlichen Bereichen bereiten die Schüler fächerverbindend auf die Arbeit mit spanischsprachigen Fachtexten in Studium und Beruf vor.“⁷³²

Die Arbeit mit realen Texten, insbesondere Fachtexten aus dem wissenschaftlichen Bereich, wird also erneut betont, hier allerdings in Verbindung mit elektronischen Medien. Da ein Korpus eine Sammlung von Texten in elektronischer Form ist und die Mehrheit von ihnen reale Texte beinhaltet, kann es im Unterricht der Qualifikationsphase durchaus eingesetzt werden, um Ziele des Lehrplans zu erreichen. Bei einigen großen Referenzkorpora wie dem CREA (siehe Kapitel 2.1.1.1.) oder dem CdE:Orig (siehe Kapitel 2.1.1.3.) kann sogar speziell nach akademischen Texten gesucht werden.

3.1.3.3. Interkulturelles Lernen und Landeskunde

In der 8. Jahrgangsstufe ist die Vermittlung der sprachlichen Vielfalt Spaniens ein Bestandteil des Lehrplans im Hinblick auf interkulturelles Lernen und Landeskunde. Die Schüler sollen demnach schon im ersten Lernjahr „autonome Regionen und Mehrsprachigkeit in Spanien“ kennenlernen.⁷³³ Wie bereits erwähnt, kann das CLUVI (siehe Kapitel 2.3.2.) dazu herangezogen werden, um die Varietäten des Katalanischen, Galicischen und Baskischen dem Kastilischen gegenüberzustellen und den Schülern anhand von ausgewählten Textausschnitten die Unterschiede zwischen den Sprachen zu verdeutlichen.

In der 9. Jahrgangsstufe, also dem zweiten Lernjahr, heißt es, „Die Schüler vertiefen ihre Kenntnisse über die spanischsprachige Welt. Die Lebensweise Gleichaltriger stellen sie ihrer eigenen vergleichend gegenüber.“ Sie sollen „Alltag und Freizeitgestaltung, insbesondere in der Welt der Jugend“ kennenlernen.⁷³⁴

⁷³² Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Spanisch (3. Fremdsprache) – Jahrgangsstufe 11/12*. Op. cit. (06.09.2017).

⁷³³ Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Spanisch (3. Fremdsprache) – Jahrgangsstufe 8*. Op. cit. (06.09.2017).

⁷³⁴ Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Spanisch (3. Fremdsprache) – Jahrgangsstufe 9*. Op. cit. (06.09.2017).

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

In der 10. Jahrgangsstufe wird dem folgend angegeben:

„Die Schüler erweitern ihre Kenntnisse über kulturelle, politische, wirtschaftliche und geographische Gegebenheiten der spanischsprachigen Welt. Durch die bewusste Wahrnehmung anderer Wertvorstellungen und Lebenswirklichkeiten, auch anhand literarischer Texte, sollen sie zu größerer Aufgeschlossenheit und einem vertieften Verständnis für andere Denkweisen sowie zur kritischen Reflexion über den eigenen Lebensentwurf gelangen.“⁷³⁵

Zudem sollen sie „exemplarische Einblicke in Kultur und Gesellschaft im spanischsprachigen Raum“ gewinnen.⁷³⁶

Für beide Jahrgangsstufen eignet sich daher der Umgang mit dem Korpus *Sacodeyl* (siehe Kapitel 2.4.3.1.), da hier spanische Jugendliche über ihre Welt, also ihren Alltag und ihre Freizeitgestaltung, aber auch über Institutionen, die Gesellschaft, die europäische Union, Zukunftspläne oder eine gesunde Lebensführung sprechen.⁷³⁷ Auch das Korpus *Backbone Spanish* (siehe Kapitel 2.4.3.2.) eignet sich für die Vermittlung dieser Inhalte, da in dessen Interviews unter anderem kulturelle, wirtschaftliche und politische Belange thematisiert werden.⁷³⁸ Während in der 9. Klasse besonders die Lebensweise Gleichaltriger in Spanien und der Vergleich ihrer eigenen mit deren Gewohnheiten von Interesse ist, rückt in der 10. Jahrgangsstufe die Sicht spanischer Jugendlicher auf deren Gesellschaft, Institutionen und Lebensführung in den Mittelpunkt. Durch die Nutzung der Interviews werden im Unterricht nicht nur Texte über diese Gegebenheiten gelesen oder gehört, sondern es wird von muttersprachlichen Jugendlichen beziehungsweise im Fall des Korpus *Backbone Spanish* von Erwachsenen verschiedener Altersgruppen darüber berichtet, welche sie in ihren eigenen Worten beschreiben.

Es kann vor allem in der 10. Klasse auch mit anderen Korpora gearbeitet werden, insbesondere wenn es sich in der Unterrichtseinheit um sprachliche Mittel oder Vergleiche zwischen verschiedenen Varietäten handelt, doch im Bereich interkulturelles Lernen und Landeskunde bringen die Korpora *Sacodeyl* und *Backbone Spanish* besondere Vorteile mit sich.

⁷³⁵ Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Spanisch (3. Fremdsprache) – Jahrgangsstufe 10*. Op. cit. (06.09.2017).

⁷³⁶ Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Spanisch (3. Fremdsprache) – Jahrgangsstufe 10*. Op. cit. (06.09.2017).

⁷³⁷ Vgl. Pérez-Paredes: *Sacodeyl. European Youth Language*. Op. cit. (18.11.2016).

⁷³⁸ Vgl. Kohn: *Backbone*. Op. cit. (15.02.2018).

3.1.3.4. Lernstrategien und Methoden selbstständigen Arbeitens

Bereits im Grundwissen der Jahrgangsstufe 8 für den Spanischunterricht ist das Wissen über die Informationsbeschaffung verankert. So sollen die Schüler mittels des Transfers der bereits erworbenen Lernstrategien und Arbeitstechniken Hilfsmittel zur Informationsbeschaffung nutzen.⁷³⁹ Hier wird insbesondere der Computer gerne von Schülern genutzt, da er aus dem Leben der meisten ohnehin nicht mehr wegzudenken ist und sehr häufig in Benutzung ist. Somit können die Schüler mittels des Computers online Datenbanken oder auch Sprachkorpora nutzen, sofern ihnen im Unterricht vermittelt wurde, wie mit diesen gearbeitet wird.

Zudem gibt das Bayerische Staatsministerium für Unterricht und Kultus für die 8. Jahrgangsstufe bezüglich der Lernstrategien und Methoden selbstständigen Arbeitens Folgendes vor:

“Die Schüler erweitern und festigen ihre Methodenkompetenz, indem sie diese im Hinblick auf die spezifischen Erfordernisse des Spanischen als dritter Fremdsprache anwenden. Sie beginnen, die Techniken und Strategien selbständig, der jeweiligen Lernsituation und Aufgabenstellung entsprechend, einzusetzen.”⁷⁴⁰

Die Schüler können mithilfe von Korpora ihre Methodenkompetenz und ihre Spanischkenntnisse selbstständig erweitern. Außerdem sollen sie „Methoden der Selbstkorrektur sowie individuell geeignete Strategien zum Erlernen und Wiederholen von Wortschatz und Grammatik anwenden“ sowie „Zusatzmaterialien zum Lehrbuch, ggf. auch neue Medien, für den eigenen Lernprozess heranziehen”.⁷⁴¹ Speziell für die Methoden der Selbstkorrektur ist die Korpusarbeit von Bedeutung. Die Schüler können anhand von Korpora ihre eigenen Sprachproduktionen korrigieren und verbessern. Sie können die Verwendung von Lemmata mithilfe der Konkordanzen vergleichen und sich Anregungen für zukünftige mündliche und schriftliche Sprachproduktionen holen. Auch die Verwendung der Grammatik kann anhand der Konkordanzen überprüft werden. Wenn sie an die Korpusarbeit gewöhnt sind, ziehen die Schüler das Korpus im besten Fall regelmäßig für den eigenen Lernprozess heran.

⁷³⁹ Vgl. Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Spanisch (3. Fremdsprache) – Jahrgangsstufe 8*. Op. cit. (06.09.2017).

⁷⁴⁰ Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Spanisch (3. Fremdsprache) – Jahrgangsstufe 8*. Op. cit. (06.09.2017).

⁷⁴¹ Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Spanisch (3. Fremdsprache) – Jahrgangsstufe 8*. Op. cit. (06.09.2017).

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

In der Qualifikationsphase heißt es im Bereich ‚Lernstrategien und Methoden selbstständigen Arbeitens‘ schließlich, die Schüler sollen ihre Fähigkeiten weiterentwickeln, indem sie „Wörterbücher, Grammatiken, Nachschlagewerke und weitere Hilfsmittel zum individuellen Lernen und Wiederholen sowie zum Erstellen und Überarbeiten eigener Texte heranziehen“.⁷⁴² Wie in Kapitel 3.3.1.2. zu sehen sein wird, wurden Korpora bereits erfolgreich zur selbstständigen Überarbeitung eigener fremdsprachlicher Texte durch Schüler herangezogen.

Der Lehrplan lässt den Schulen und insbesondere den Lehrern viel Freiraum in Bezug auf die Gestaltung des Sprachunterrichts in den verschiedenen Jahrgangsstufen. Er legt generelle Richtlinien und Ziele fest, beschreibt jedoch nicht die Art und Weise, wie diese erreicht werden sollen. Somit kann die einzelne Lehrkraft durchaus neben der Lehrbucharbeit mit Texten aus dem Internet, online verfügbaren oder selbst erstellten Korpora arbeiten, wodurch der Unterricht abwechslungsreicher und interessanter wird.

3.2. Motive für die Verwendung von Korpora im Unterricht

Im Hinblick auf den lerntheoretischen Hintergrund haben sich in den bisherigen Ausführungen drei für den modernen Fremdsprachenunterricht wichtige Kriterien herauskristallisiert, welche durch die Arbeit mit Korpora unterstützt werden können: die sogenannte Authentizität, die kommunikative Kompetenz sowie die Lernerautonomie. Diese Begriffe sollen daher im Folgenden genauer betrachtet werden, um zu klären, was genau darunter zu verstehen ist, insbesondere vor dem Hintergrund des Fremdsprachenunterrichts.

3.2.1. Authentisch, genuin, real oder realistisch – der Begriff der sogenannten Authentizität

Authentizität ist eines der obersten Ziele der Fremdsprachendidaktik und als solches in aller Munde, doch es stellt sich die Frage, was der Begriff ‚authentisch‘ konkret bedeutet. Bludau gibt an, der Begriff der Authentizität sei extrem mythologisiert worden

⁷⁴² Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Spanisch (3. Fremdsprache) – Jahrgangsstufe 11/12*. Op. cit. (06.09.2017).

und schwebe als Zauberwort über den Lehrplänen.⁷⁴³ Rüschoff/Wolff bezeichnen ihn als „schillernden Begriff in der fremdsprachendidaktischen Diskussion“.⁷⁴⁴

Gemäß des *Wörterbuchs für Erziehung und Unterricht* von Köck bezeichnet der Begriff Authentizität „die Echtheit, Glaubwürdigkeit oder Zuverlässigkeit einer Sache, einer Person mit ihren Verhaltensweisen oder eines Mediums“.⁷⁴⁵ Als Beispiel führt Köck ein Videoband an, welches nur dann ‚authentisch‘ sei, wenn es nach der Aufzeichnung weder inhaltlich, durch spätere Einblendungen oder nachträgliches Überspielen, noch technisch verändert wurde.⁷⁴⁶ Dennoch bleibt die Frage offen, was Authentizität im Hinblick auf den Unterricht genau bedeutet.

In die Fremdsprachendidaktik hält der Begriff der Authentizität im Gefolge der *Kommunikativen Wende* am Ende der 1970er Jahre Einzug (siehe Kapitel 3.1.1.). Während der 1980er Jahre gewinnt er immer mehr an Bedeutung und entwickelt sich in den 1990er Jahren zum Schlüsselbegriff.⁷⁴⁷

Widdowson differenziert auf diesem Gebiet bereits im Jahr 1979 zwischen genuinen, also echten Materialien einerseits, und einer ‚authentischen‘ Verwendung derselben andererseits. Er setzt also den Umgang mit muttersprachlichem Material ins Zentrum der Diskussion. Widdowson sieht den Begriff der Authentizität als einen relationalen, welcher sich aus der Beziehung zwischen dem Text und dem Leser ergibt. Als Beispiel führt er einen Zeitungsartikel an, welcher für gewöhnlich gelesen wird, um sich über das Thema desselben zu informieren. Wird nun ein Schüler im Fremdsprachenunterricht dazu angehalten, denselben Text zu lesen, um daraus sprachliche Informationen zu erhalten, so ist die Beziehung eine ganz andere, keine ‚authentische‘ mehr. Der Zeitungsartikel an sich hat sich allerdings nicht verändert, er ist ein genuiner Text, da er nicht eigens für Lernzwecke erstellt, sondern für muttersprachliche Rezipienten geschrieben wurde. Die Eigenschaft der Genuität wohnt jedem Text inne und ist nach Widdowsons Meinung nicht von der Beziehung zwischen Text und Rezipienten abhängig. Gegenteilig verhält es sich mit der Authentizität. Diese

⁷⁴³ Vgl. Bludau, Michael (1996): „What is ‘authentic’?“. In: *Zielsprache Englisch*, 1996, 26 (2), S. 11-13.

⁷⁴⁴ Rüschoff, Bernd/ Wolff, Dieter (1999): *Fremdsprachenlernen in der Wissensgesellschaft. Zum Einsatz der neuen Technologien in Schule und Unterricht*. Ismaning: Hueber, S. 61.

⁷⁴⁵ Köck (2008): op. cit., S. 38.

⁷⁴⁶ Vgl. Köck (2008): op. cit., S. 38.

⁷⁴⁷ Vgl. Decke-Cornill, Helene (2004): „Die Kategorie der Authentizität im mediendidaktischen Diskurs der Fremdsprachendidaktik“. In: Bosenius, P./ Donnerstag, J. (Hrsg.) (2004): *Interaktive Medien und Fremdsprachenlernen*. Frankfurt am Main: Lang, S. 17-18.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

ist keine dem Text inhärente Eigenschaft, sondern sie ergibt sich aus der Verwendung desselben durch Produzenten und Rezipienten. Ein Text kann seiner Meinung nach nur als ‚authentisch‘ bezeichnet werden, wenn er mit derjenigen Intention gelesen wird, mit der er erstellt wurde. Daher handelt es sich bei den gängigen nicht-didaktisierten Materialien um genuine, nicht aber um ‚authentische‘ Materialien. Schüler und Lehrer müssen genuine Texte also in der jeweiligen Unterrichtssituation erst authentisieren. Hierfür müssten die Lerner nach Meinung von Widdowson mit den Texten so umgehen, wie sie beim Diskurs in der Muttersprache verwendet würden und nicht mit anderen Zielsetzungen. Ein Vorschlag für die Authentisierung des Unterrichts wäre daher die Auswahl von Texten, welche für die Schüler interessant sind und daher wie muttersprachliche Texte aus echtem Interesse gelesen werden, um daraus kulturelle und keine sprachlichen Informationen zu entnehmen.⁷⁴⁸

Für Widdowson kann also lediglich die Situation, in welcher der Text gelesen wird, ‚authentisch‘ sein, ein Text an sich ist seiner Meinung nach genuin. Die Problematik der Schaffung ‚authentischer‘ Situationen im Klassenzimmer sieht auch Pearse in seinem Aufsatz „Realistic and Real English in the Classroom“ aus dem Jahr 1983. Er verwendet die Begriffe jedoch auf eine andere Art und Weise. Texte, ebenso wie Gegenstände aus dem zielsprachigen Land stuft er eindeutig als ‚authentisch‘ ein, wenn der Lehrer diese jedoch im Unterricht verwendet, so ist ihr Gebrauch höchstens realistisch, keineswegs real. Seiner Meinung nach spielt das *setting* eine bedeutende Rolle im fremdsprachlichen Klassenzimmer. Das *setting* des Fremdspracherwerbs – das Klassenzimmer – steht den zahlreichen *settings*, für die gelernt werden soll, gegenüber. Die Welt außerhalb des Klassenzimmers muss daher in dasselbe gebracht werden, was allerdings nur bedingt möglich ist.⁷⁴⁹

Zunächst unterscheidet Pearse die Begriffe *real* und *realistic*. Er bemängelt, dass Lehrer oft denken, sie präsentieren ihren Schülern reales Englisch, während sie ihnen jedoch nur realistische Übungen vorsetzen. Er bemerkt treffend: „Real English is when you use language to say something. Realistic English is when you say something to use language.“⁷⁵⁰ Er veranschaulicht diesen Sachverhalt mit der Frage nach der Uhrzeit. Im

⁷⁴⁸ Vgl. Widdowson, Henry G. (1979): *Teaching Language as Communication*. Oxford: University Press, S. 80-81.

⁷⁴⁹ Vgl. Pearse, Roy (1983): „Realistic and Real English in the Classroom“. In: *English Teaching Forum*, 21/3, S. 19-20.

⁷⁵⁰ Pearse (1983): op. cit., S. 20.

ersten Fall fragt ein Schüler danach, weil er vielleicht zu spät zu seinem Zahnarzttermin kommt oder noch schnell Milch kaufen will, bevor die Geschäfte schließen. Im zweiten Fall fragt er nach der Uhrzeit, weil dies das Thema der Stunde ist und der Lehrer die Zeiger der Uhr auf seinem Schreibtisch verstellt, um den Schüler nach der Zeit zu fragen. Hier ist also die Sprache von höchster Bedeutung und nicht die Motivation. Die Situation ist realistisch, aber nicht real. Obgleich das sprachliche Material von Muttersprachlern produziert wurde, ist das *setting* im Klassenzimmer doch immer nur realistisch, wenn nicht sogar artifiziell.⁷⁵¹

Edelhoff differenziert zwei Jahre später in seinem Aufsatz „Authentizität im Fremdsprachenunterricht“ ebenfalls zwischen der sogenannten sprachlich-linguistischen und der pädagogisch-situativen Authentizität. Die erste sei vor allem für die Auswahl und Vorbereitung von Texten für den Unterricht von Bedeutung. Sie befasse sich mit der Qualität der Textauswahl, -bearbeitung und -anordnung, da die Repräsentationen der Fremdsprache in der Schule der Fremdsprachenbegegnung und -anwendung im Leben möglichst nahe sein sollten. Die pädagogisch-situative Authentizität beziehe sich vor allem auf die unterrichtlich-interaktive Seite des Lernprozesses. Er benennt die beiden Arten der Authentizität auch als die materielle Seite des Curriculums einerseits, und die pädagogisch-methodische Seite desselben andererseits.⁷⁵² Edelhoff ist der Meinung, dass die Authentizität der Texte eine wichtige Rolle im Fremdsprachenunterricht spielt, da ein Text eben nicht nur zu erlernende sprachliche Formen enthalten, sondern auch Inhalte vermitteln sollte. Es können zwar nicht alle Texte in ihrer ursprünglichen Form verwendet werden, doch die Forderung nach Authentizität darf seiner Meinung nach nie außer Acht gelassen werden. In Bezug auf die pädagogisch-situative Authentizität bemängelt er vor allem die damals in Europa vorherrschende, aus dem Latein- und Griechischunterricht stammende Übersetzungsmethode.⁷⁵³

Auch Rüschoff/Wolff unterscheiden diese beiden Arten der Authentizität. Als Beispiele für ‚authentische‘ Interaktionen im Unterricht werden die selbstständige Suche nach Beispielen für bestimmte Strukturen oder nach verständlichen Kontexten für zu

⁷⁵¹ Vgl. Pearse (1983): op. cit., S. 20.

⁷⁵² Vgl. Edelhoff, Christoph (1985): „Authentizität im Fremdsprachenunterricht“. In: Edelhoff, C. (Hrsg.) (1985): *Authentische Texte im Deutschunterricht*. München: Hueber, S. 7.

⁷⁵³ Vgl. Edelhoff (1985): op. cit., S. 8, 24.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

erlernendes Vokabular angegeben sowie das eigenständige Erschließen von Zusammenhängen.⁷⁵⁴ Derartige selbstständige Suchen nach Belegen für sprachliche Strukturen können beispielsweise anhand der Korpora durchgeführt werden, wie die Anwendungsbereiche im vierten Kapitel zur Nutzung von Korpora in der fremdsprachlichen Unterrichtspraxis zeigen werden.

Ebenfalls im Jahr 1985 zählt Breen in seinem Aufsatz „Authenticity in the Language Classroom“ zwei weitere Arten der Authentizität auf:

1. Die Authentizität der Texte, die als Input für die Lerner dienen.
2. Die Authentizität der Interpretation solcher Texte durch den Lerner selbst.
3. Die Authentizität der Aufgaben, die dem Fremdsprachenlernen dienlich sind.
4. Die Authentizität der konkreten sozialen Situation im Klassenzimmer.

Er stellt fest, dass sich die Debatte über Authentizität in der Vergangenheit hauptsächlich mit der ersten Art beschäftigte. Allerdings existierten für jede der vier genannten Arten – Inhalt, Lerner, Lernen und Klassenzimmer – unterschiedliche Kriterien für Authentizität, welche es im Unterricht in Balance zu halten gelte.⁷⁵⁵

In Bezug auf die Authentizität von Texten konstatiert Breen, dass diese immer in Zusammenhang mit der Authentizität der Interpretation derselben durch den Lerner steht. Er verweist bei dieser Art der Authentizität auch auf Widdowsons Unterscheidung zwischen der Texten innewohnenden Genuität und der Authentizität der Interaktion zwischen Lerner und Text, welche bereits erwähnt wurde. Nach Breens Meinung kann ein für Muttersprachler geschriebener Text nicht ohne Weiteres für den Unterricht mit Fremdsprachenlernern eingesetzt werden, da diese den Text stets vor dem Hintergrund der Vermittlung bestimmter lexikalischer und grammatikalischer Strukturen sehen und denselben aufgrund ihres Referenzrahmens immer anders interpretieren als Muttersprachler. Lerner sind daran gewöhnt, dass die ihnen in Lehrbüchern vermittelten Texte primär metakommunikative Ziele verfolgen, also darauf abzielen, Funktionen der zu erlernenden Sprache zu vermitteln oder, wie Breen es nennt, ihnen den *Code* der Sprache nahezubringen. Im Gegensatz dazu nutzt ein Autor, der Texte für Muttersprachler schreibt, die Strukturen der Sprache, also eben diesen *Code*, um auf

⁷⁵⁴ Vgl. Rüschhoff/ Wolff (1999): op. cit., S. 61-62.

⁷⁵⁵ Vgl. Breen, Michael P. (1985): „Authenticity in the Language Classroom“. In: *Applied Linguistics*, Vol. 6, No. 1, S. 61.

möglichst effektive Art und Weise eine bestimmte Bedeutung zu übermitteln – er verfolgt also kommunikative Ziele.⁷⁵⁶ Auch für Little/Devitt/Singleton stellt der kommunikative Zweck das Hauptkriterium der Definition ‚authentischer‘ Texte dar.⁷⁵⁷ Somit gelten die traditionellen Lehrbuchtexte nicht als ‚authentisch‘, da ein solcher Text kein kommunikatives Ziel in einer Sprachgemeinschaft verfolgt, sondern lediglich dem Zweck des Fremdsprachenlernens dient. Lerner sind damit immer weiter von der eigentlichen Intention des Texts entfernt als Muttersprachler, sie nähern sich dieser allerdings, je mehr sie über die Zielsprache erfahren. Mit der Frage nach einem ‚authentischen‘ Text stellt sich also auch immer die Frage danach, für wen der jeweilige Text ‚authentisch‘ ist, und damit für den Lehrer auch die Frage nach den möglichen ‚authentischen‘ Zugängen an den Text. Im Unterricht sollen also Kriterien vermittelt werden, welche die Schüler nach und nach zu einer Interpretation führen, die derjenigen von Muttersprachlern sehr nahe kommt.⁷⁵⁸ Zudem bemängelt Kramersch im Hinblick auf die Verwendung von Texten aus dem fremdsprachlichen Ausland, dass das Verständnis von Texten, die für Muttersprachler geschrieben wurden, teilweise das Hintergrundwissen eines dauerhaft im Ausland lebenden Mitbürgers sowie einen Wechsel der Perspektive erfordert und diese Texte daher nicht unbedingt eine ‚authentische‘ Lesart ermöglichen.⁷⁵⁹

Bezüglich der dritten Art, der Authentizität der Aufgaben, die dem Fremdsprachenlernen dienlich sind, unterscheidet Breen zwischen dem Verhalten im Sprachgebrauch einerseits und dem im Sprachlernen andererseits, wobei die Authentizität des einen nicht zwangsläufig die des anderen bedeutet. Ganz im Gegenteil, seiner Meinung nach ist es durchaus möglich, dass Aufgabenformen zwar nicht ‚authentisch‘ in Bezug auf den Sprachgebrauch sind, aber ‚authentisch‘ bezüglich des Sprachlernens. Als gutes Beispiel hierfür führt Breen eine Aufgabe an, bei welcher Schüler einen Text lesen und alle für sie schwer zu interpretierenden Passagen unterstreichen sollen. Daraufhin werden sie angehalten, in Paaren zusammen zu arbeiten und zu versuchen, gemeinsam so viel wie möglich zu verstehen. Die noch unklaren Passagen sollen dann eingekreist werden. Diese Aufgabenstellung leitet die Schüler an, auf ‚authentische‘ Art und Weise miteinander zu kommunizieren und führt gleichzeitig zur ‚authentischen‘ Erfahrung

⁷⁵⁶ Vgl. Breen (1985): op. cit., S. 61-63.

⁷⁵⁷ Vgl. Little, David/ Devitt, Seán/ Singleton, David M. (1994): *Learning foreign languages from authentic texts. Theory and practice*. Dublin: Authentik Language Learning Resources, S. 23.

⁷⁵⁸ Vgl. Breen (1985): op. cit., S. 61-63.

⁷⁵⁹ Vgl. Kramersch (1993): op. cit., S. 185-189.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

neuer Lerninhalte. Die Schüler sollen ihre Ideen und Meinungen *kommunizieren* und über die Sprache selbst, die Probleme und Lösungen im Sprachlernprozess *metakommunizieren*. Als weiteres Beispiel beschreibt Breen eine Aufgabenstellung, bei der die Schüler nach einem vorgegebenen Schema auf Aussagen reagieren sollen, indem sie diesen entweder zustimmen (*So do I.*) oder diese verneinen (*Neither do I.*). Derartige Gespräche sind bei der Kommunikation in der zielsprachlichen Kultur äußerst unwahrscheinlich, haben allerdings eine Berechtigung in Bezug auf das Sprachlernen. Daher räumt Breen zum Schluss der Bewertung dieser Aufgabe ein „such apparently inauthentic language-using behaviour might be authentic language *learning* behaviour“.⁷⁶⁰

Im Bereich der Authentizität der konkreten sozialen Situation im Klassenzimmer rät Breen, diese spezielle soziale Situation zu nutzen, indem die Schüler als sie selbst über die Zielsprache und ihre Probleme beim Erlernen derselben sprechen können. Das Klassenzimmer soll nicht zu etwas gemacht werden, was es nicht ist, und gleichzeitig soll die Rolle der Schüler nicht verändert werden, da eine solche Veränderung die Authentizität des Klassenzimmers einschränken würde.⁷⁶¹

Die Reichweite des Authentizitätsbegriffs nahm über die Jahre also immer mehr zu. Zunächst bezog sich der Begriff lediglich auf Unterrichtsmaterialien und deren Herkunft aus zielkulturellen Kontexten, später dann auch auf die Interaktionen im Unterricht und das gesamte Unterrichtsgeschehen. Schließlich wurde er auf die am Sprachunterricht beteiligten Personen und das Handeln der Schülerinnen und Schüler untereinander erweitert.

Die Authentizität, vor allem diejenige der Unterrichtsmaterialien, ist demnach ein von vielen gefordertes, jedoch vor allem bei Lernern im ersten Lernjahr schwer zu erreichendes Kriterium. Einige Autoren sprechen sich deshalb gegen die Verwendung ‚authentischer‘ Materialien im Unterricht aus.⁷⁶² Sie sind der Meinung, die Texte seien zu komplex für Fremdsprachenlerner und überforderten sie aufgrund des zu hohen Schwierigkeitsgrads. Diese Überforderung könnte zu Demotivation und Frustration seitens der Schüler führen. Aus diesem Grund haben einige Didaktiker neue Konzepte in

⁷⁶⁰ Breen (1985): op. cit., S. 65.

⁷⁶¹ Vgl. Breen (1985): op. cit., S. 66-68.

⁷⁶² Vgl. Bludau (1996): op. cit.; Conolly (1996): op. cit.; Krashen, Stephen D. (1996): „Not pedagogical or authentic, but interesting and comprehensible...“. In: *Zielsprache Englisch*, 1996, 26 (2), S. 22-25.

Bezug auf die Authentizität der Texte entwickelt. Bludau spricht beispielsweise von der *near-authenticity* und definiert diese wie folgt:

„Der Fremdsprachenlehrer/Lehrbuchautor usw. sollte zunächst den zu vermittelnden Stoff (= die Fremdsprache) in eine Form bringen, die den Voraussetzungen, Bedingungen und Bedürfnissen seiner jeweiligen Lerngruppe entspricht und angemessen ist. Ein solches – unter sprachdidaktischen Gesichtspunkten aufbereitetes – Material muß [sic!] aber nach Sprache und Sache für eine repräsentative Auswahl von native speakers akzeptabel sein.“⁷⁶³

Lademann schlägt hierfür den Begriff *pedagogically authentic* oder ‚angemessen authentisch‘ vor, wobei für ihn vor allem auf das Interesse der Schüler am Text geachtet werden sollte.⁷⁶⁴ Padrós/Biechele stützen sich auf den Begriff der ‚gemäßigten Authentizität‘, welche beispielsweise vorliegt, wenn die verwendeten Texte bezüglich ihrer Sprache und Form der entsprechenden Textsorte angemessen sind. Vor allem für Anfänger müssen die Texte ihrer Meinung nach so gewählt werden, dass es keine unüberwindbaren Verständnishürden gibt und das sprachliche Material in das bereits vorhandene Wissen der Lernenden integrierbar ist.⁷⁶⁵

Decke-Cornill ist der Meinung, dass keine der genannten Arten von Authentizität im Unterricht erreicht werden kann und plädiert dafür, das Konzept der Authentizität nicht über alles zu stellen, sondern anstatt dessen das Konzept des guten Fremdsprachenunterrichts. Sie sieht in der schulischen Situation an sich allerdings eine ganz eigene Authentizität und bezeichnet diese als ‚konstruierte Echtheit‘ des fremdsprachlichen Klassenzimmers.⁷⁶⁶ Dieser Ansicht schließt sich auch Rössler an und beschreibt die möglichst ‚authentische‘ Gestaltung didaktisierter Materialien als ‚simulierte Authentizität‘. Sie schreibt derartigen Materialien das Ziel zu, die Schüler möglichst effektiv auf ‚authentische‘ Kommunikationssituationen vorzubereiten.⁷⁶⁷

Als problematisch stellt sich also nicht nur das Konzept der sogenannten Authentizität dar, sondern auch die Begrifflichkeit. Es gibt unterschiedliche Arten der Authentizität sowie zahlreiche Namen, die teils mit derselben teils mit einer völlig

⁷⁶³ Bludau (1996): op. cit., S. 13.

⁷⁶⁴ Vgl. Lademann, Norbert (1997): „Authentizität, Text und Sprechintention“. In: Siebold, J. (Hrsg.) (1997): *Sprache und Medien im Fremdsprachenunterricht. Beiträge des 3. Mediendidaktischen Kolloquiums Oktober 1996 in Rostock*. Rostock: Universitätsdruckerei Rostock, S. 45.

⁷⁶⁵ Vgl. Padrós, Alicia/ Biechele, Markus (2003): *Didaktik der Landeskunde*. Berlin/ München: Langenscheidt, S. 42.

⁷⁶⁶ Vgl. Decke-Cornill (2004): op. cit., S. 25.

⁷⁶⁷ Vgl. Rössler, Andrea (2010): „Zur Frage der Authentizität von Lehrwerkdialogen für den Französischunterricht“. In: Frings, M./ Leitzke-Ungerer, E. (Hrsg.) (2010): *Authentizität im Unterricht romanischer Sprachen*. Stuttgart: ibidem-Verlag, S. 29-30.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

unterschiedlichen Bedeutung verwendet werden. Aufgrund der Beschäftigung mit Korpora und daraus entnommenen Texten spielt für die vorliegende Arbeit vor allem die Authentizität der Unterrichtsmaterialien eine Rolle. Die Bezeichnung ‚authentisch‘ ist jedoch nicht eindeutig, wie in diesem Kapitel zu sehen war, und aus diesem Grund soll für derartige Texte der Begriff ‚real‘ verwendet werden. Dem Begriff der realen Texte soll die Authentizitätsdefinition von Rössler zugrunde gelegt werden, welche sich der vorherrschenden Meinung über die Authentizität von Materialien anschließt, wenn sie diese wie folgt definiert:

„Unter authentischen Materialien versteht man Texte aller Art, die nicht speziell für den Einsatz im Fremdsprachenunterricht entwickelt worden sind, sondern von Mitgliedern einer bestimmten Sprachgemeinschaft für andere Mitglieder derselben Sprachgemeinschaft zwecks Erfüllung einer kommunikativen Absicht und sozialen Handlung realisiert wurden.“⁷⁶⁸

Das Konzept realer Texte soll in Anlehnung an Rössler (2010) also wie folgt definiert werden: Ein realer Text ist ein solcher, welcher von einem muttersprachlichen Produzenten an ein muttersprachliches Publikum mit einer bestimmten Kommunikationsintention, jedoch nicht speziell für den Fremdsprachenunterricht, geschrieben wurde. Somit wird in der vorliegenden Arbeit von nun an der Begriff ‚real‘ verwendet werden, weil der Begriff ‚authentisch‘ nicht klar definiert werden kann und durch die fremdsprachendidaktische Diskussion vorbelastet ist.

Es soll jedoch bereits jetzt klargestellt werden, dass diese Arbeit nicht den vollständigen Ersatz didaktisierter Unterrichtsmaterialien durch reale Texte propagieren, sondern eine Erweiterung der traditionellen Lehrbuchtexte um reale Texte vorantreiben will. Auch die didaktische Aufbereitung der aus Korpora gewonnenen Sprachdaten stellt keine Hürde dar, unter der Voraussetzung, dass die dargebotenen Situationen und Aufgabenstellungen realistisch sind und die Schüler im spanischsprachigen Ausland durchaus in solche verwickelt werden können. Schiffler beispielsweise ist der Meinung, ein Originaltext kann für den Unterricht gekürzt oder durch eine zum Gesamtzusammenhang des Schulbuchs passende Einleitung ergänzt werden. Auch der Austausch vulgärer Ausdrücke ist gerechtfertigt.⁷⁶⁹ Lernmaterialien, die vollständig aus realen, nicht-didaktisierten Texten bestehen, würden die Schüler richtiggehend

⁷⁶⁸ Rössler (2010): op. cit., S. 29.

⁷⁶⁹ Vgl. Schiffler, Ludger (2012): *Effektiver Fremdsprachenunterricht*. Tübingen: Narr Francke Attempto, S. 33.

‚erschlagen‘, so sollen die aus Korpora entnommenen Texte vielmehr als Zusatz für die bisherigen Materialien dienen und mehr Vielfalt ins Klassenzimmer bringen.

Die Verwendung realer Materialien ist auch immer vom Lernstadium der Schüler abhängig. Während Abiturienten grundsätzlich mit jeglichen realen Daten konfrontiert werden können, muss bei Lernern im ersten und auch zu Beginn des zweiten Lernjahrs noch verstärkt darauf geachtet werden, dass diese die Dokumente auch verstehen. Allerdings können sogar zu Beginn des Fremdsprachenlernens bereits Auszüge aus Korpora eingesetzt werden, zum Beispiel in Form einfacher Konkordanzen oder Kollokationen.

Maria Ciesielska-Ciupek spricht der Arbeit mit Korpora aufgrund der realen Materialien auch einen positiven Einfluss in Bezug auf die kommunikative Kompetenz zu. Sie beschreibt die kommunikative Kompetenz als Festigung struktureller, sozialer und strategischer Kompetenz. Eine gute kommunikative Kompetenz beinhaltet demnach guten Sprachgebrauch außerhalb des Klassenzimmers in natürlicher sprachlicher Umgebung, und die Arbeit mit Korpora ermöglicht ebenso wie die Verwendung des Internets unmittelbare sprachliche Kommunikation mit anderen Sprechern.⁷⁷⁰

3.2.2. Der Begriff der kommunikativen Kompetenz

Ziel des Fremdsprachenunterrichts ist die Ausbildung einer sprachlichen und interkulturellen Handlungsfähigkeit in Alltagssituationen im Zielsprachenland. Die kommunikative Kompetenz ist also eine Forderung, die an den modernen Fremdsprachenunterricht herangetragen wird. Es stellt sich dabei allerdings die Frage, was unter diesem Begriff zu verstehen ist. In Bußmanns *Lexikon der Sprachwissenschaft* wird die kommunikative Kompetenz als der Grundbegriff eines pragmalinguistisch orientierten Modells sprachlicher Kommunikation gesehen und sie wird definiert als:

„allgemeine Sprachfähigkeit von Individuen, die in der Lage sind, im Einklang mit wechselnden situativen und normativen Bedingungen psychischer, sozialer und linguistischer Natur miteinander zu kommunizieren, wobei Sprechen als symbolvermitteltes Handeln (= Interaktion) verstanden wird.“⁷⁷¹

⁷⁷⁰ Vgl. Ciesielska-Ciupek, Maria (2003): „Teaching with the internet and corpus materials. Preparation of the ELT materials using the Internet and corpus resources“. In: Lewandowska-Tomaszczyk, B. (Hrsg.) (2003): *PALC 2001: Practical Applications in Language Corpora*. Frankfurt am Main: Lang, S. 523-524.

⁷⁷¹ Bußmann (2008): op. cit., S. 348.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

Weiter sei die kommunikative Kompetenz in Deutschland ein von der kommunikationsorientierten Soziologie postuliertes, idealisiertes Modell der sprachlichen Kommunikation als Teil einer Handlungstheorie, das sich vor allem auf die pragmatischen Komponenten der Konstituierung von möglichen Redesituationen und auf die Fähigkeit zur situationsangemessenen Herstellung sozialer Beziehungen durch sprachliche Interaktionen bezieht.⁷⁷²

Im *Wörterbuch für Erziehung und Unterricht* von Peter Köck findet sich unter dem Schlagwort *kommunikative Kompetenz* folgende Definition:

„Kommunikative Kompetenz [...] ist zugleich Voraussetzung und lebenslanges Lernziel des vor allem von J. Habermas (geb. 1929) beschriebenen Idealkonstrukts eines kommunikativen Handelns. Dieses liegt vor, wenn in gemeinsamen Lebenssituationen ein kooperativer Deutungsprozess vorgenommen wird.

Die Handlung wird an der Verständigung orientiert, d.h., wechselseitige Geltungsansprüche werden intersubjektiv anerkannt, die Handlungspläne der Beteiligten werden durch verständigungsorientierte Benutzung der Sprache koordiniert (im Gegensatz zur wechselseitigen Beeinflussung beim strategischen Handeln).“⁷⁷³

Beiden Definitionen ist gemeinsam, dass die Gesprächspartner sich in der jeweiligen Situation mittels der Sprache verständigen und die Bedeutung des jeweils Gesagten aushandeln müssen. Bevor es jedoch zur Interpretation des Dialogs kommen kann, müssen die Beteiligten über die gleichen sprachlichen Mittel verfügen, was für den Fremdsprachenunterricht konkret bedeutet, dass der Lerner den zu vermittelnden Sachverhalt sprachlich so nahe am muttersprachlichen Original wie möglich ausdrücken können sollte. Hier kommt die Nutzung von Korpora ins Spiel. Allein die didaktisierten Lehrbuchdialoge verhelfen dem Lerner noch nicht zum benötigten Sprachvermögen, doch die Beschäftigung mit muttersprachlichem Material, wie es die Korpora bereitstellen, führen ihn immer näher an das Original heran.

Hans Eberhard Piepho übertrug die Theorie von Habermas schließlich auf die Fremdsprachendidaktik. Kommunikative Kompetenz wurde daraufhin beschrieben als eine Fähigkeit, die sich erlernen lässt und unter anderem in die vier Fertigkeiten Hören, Sprechen, Lesen und Schreiben untergliedert ist.⁷⁷⁴ Er sieht die kommunikative Kompetenz vor dem Hintergrund aller wissenschaftlichen Richtungen, durch die diese

⁷⁷² Vgl. Bußmann (2008): op. cit., S. 348.

⁷⁷³ Köck (2008): op. cit., S. 253.

⁷⁷⁴ Vgl. Padrós/ Biechele (2003): op. cit., S. 42.

beeinflusst wurde, wie beispielsweise die Sozialphilosophie, die Gesellschaftswissenschaften, die Kommunikationstheorie sowie die Psycholinguistik, und definiert sie wie folgt:

„*Kommunikative Kompetenz* ist die umfassende Fähigkeit eines Sprechers, kommunikativ zu handeln und sich im Diskurs zu äußern, d. h. die Bedeutung und die Absicht in oder hinter einer Äußerung bzw. einem Text aufzufassen und eigene Absichten in wirksamer Weise sprachlich auszudrücken.“⁷⁷⁵

Weiter definiert Piepho den Begriff des kommunikativen Handelns als „die Fähigkeit, die konventionellen Sprachmittel gezielt zu bestimmten Zwecken der gesellschaftlichen Regelung des Miteinanders einzusetzen“. ⁷⁷⁶ Im Sinne der Pragmalinguistik steht nun nicht mehr die Lernbarkeit von Texten und Dialogen im Mittelpunkt, sondern die Präzisierung sprachlicher Mitteilungsleistungen. So müssen Schüler unter anderem lernen, den Lehrer höflich zu unterbrechen, mitzuteilen, wenn sie etwas nicht verstanden haben, oder um eine Wiederholung zu bitten.⁷⁷⁷

Diese neue Betrachtung des Sprachlernprozesses hat unter anderem folgende Konsequenzen:

- Sprachliche Äußerungen werden nach *kommunikativen Funktionen* dargestellt und nicht mehr nach formal-grammatischen Kategorien: *Einen Wunsch äußern* ist eine kommunikative Funktion, die durch unterschiedliche Redemittel realisiert werden kann – etwa durch *Ich möchte gern ...*, aber auch durch den Konjunktiv II *Ich hätte gern...*
- Die Progression orientiert sich an Kommunikationsfunktionen, nicht an Grammatik: Deshalb kann die Äußerung *Ich hätte gern...* beispielsweise auch schon gelernt werden, wenn der Konjunktiv II noch nicht Lernstoff war.
- Die kommunikative Angemessenheit ist von Bedeutung: Lernende sollen nicht nur korrekt sprechen, sondern wissen, welche Sprachebene angemessen ist. Die Äußerung *Vielen Dank* ist einem Freund gegenüber angemessener als etwa *Ich bin dir sehr verbunden*.
- Das Wissen über Sprache erhält einen hohen Stellenwert.⁷⁷⁸

⁷⁷⁵ Piepho, Hans-Eberhard (1974): *Kommunikative Kompetenz, Pragmalinguistik und Ansätze zur Neubestimmung in der Lernzielbestimmung im Fremdsprachenunterricht*. Düsseldorf: Winterscheidt, S. 8.

⁷⁷⁶ Piepho (1974): op. cit., S. 8.

⁷⁷⁷ Vgl. Piepho (1974): op. cit., S. 10.

⁷⁷⁸ Vgl. Padrós/ Biechele (2003): op. cit., S. 43.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

Diese Faktoren sind unter dem Begriff der *kommunikativen Wende* bekannt geworden, welche im Fremdsprachenunterricht in Zusammenhang mit einem entscheidenden Wendepunkt in der Didaktik in Deutschland steht (siehe Kapitel 3.1.1.): Bis dahin sollte auch der Fremdsprachenunterricht im Sinne der humanistischen Bildungsansprüche dazu dienen, den Geist und die Persönlichkeit der Schüler zu formen. Seit Mitte der 1960er-Jahre werden nun zunehmend pragmatische Lernziele angestrebt, welche sich an den aktuellen Erfordernissen orientieren und sich in den übergeordneten Lernzielen *kommunikative Kompetenz* und *Emanzipation* niederschlagen. Der Erwerb von Kommunikationsfähigkeit wird als Teil der Persönlichkeitsbildung betrachtet und umfasst neben der individuellen Sprachkompetenz auch die Fähigkeit zur sozialen Interaktion. In diesem Sinne sollte zudem die Isolierung von sprachlichem wie auch landeskundlichem Lernen aufgehoben werden.⁷⁷⁹

Die kommunikative Kompetenz besteht demnach aus drei Bereichen: die linguistische Kompetenz, die soziolinguistische Kompetenz und das Weltwissen. Zunächst ist die linguistische Kompetenz, welche lexikalisches, grammatisches, semantisches und phonetisches Wissen beinhaltet, ausschlaggebend für die Aneignung einer Fremdsprache. Die soziolinguistische Kompetenz, welche dem Lerner dazu verhilft, sich situationsadäquat äußern zu können, umfasst die angemessene Kenntnis der sozialen Dimension beim Sprachgebrauch sowie pragmatische Fähigkeiten. Der dritte Bereich ist schließlich das Weltwissen, welches zusammen mit dem soziokulturellen Wissen die Grundlage für erfolgreiche Kommunikation darstellt. Im Verlauf des Lernprozesses muss das eigene Weltwissen immer wieder hinterfragt und erweitert werden.⁷⁸⁰

Vor diesem Hintergrund wird klar, auf welche Weise die Korpuslinguistik der Fremdsprachendidaktik nützlich sein kann: So sollte die Auswahl dessen, was zu welchem Zeitpunkt gelernt wird, doch anhand des tatsächlichen Gebrauchs der Sprache durch Muttersprachler entschieden werden. Ein Beispiel hierfür, welches bereits seit geraumer Zeit im Spanischunterricht praktiziert wird, ist die Verwendung der Phrase *me gustaría*. Dieser Ausdruck wird bereits im ersten Lernjahr des Spanischunterrichts vermittelt, obwohl die detaillierte Auseinandersetzung mit dem *Condicional* erst im zweiten beziehungsweise dritten Lernjahr erfolgt. Mithilfe der Korpora ist es möglich,

⁷⁷⁹ Vgl. Padrós/ Biechele (2003): op. cit., S. 43.

⁷⁸⁰ Vgl. Grünewald/ Küster (2009): op. cit., S. 122.

frequente Grammatik- und Vokabeleinheiten wie diese ausfindig zu machen, welche dann, vielleicht früher als die grammatische Progression des derzeitigen Lehrplans und der Lehrwerke es vorsehen, in den Unterricht integriert werden können.

Schlickau sieht in der Verwendung der Korpora vor allem Vorteile für die Aneignung von Muttersprachlern ähnlichem Handeln und Verhalten. Als Beispiel führt er den alltäglichen Kontakt mit einem Postbediensteten aufgrund der Übergabe eines Einschreibens an, welche ganz ohne Versprachlichung der Intentionen der beteiligten Personen ablaufen kann. Jedoch wird bei diesem Beispiel klar, dass ausländische Mitbürger, um derartige Handlungen richtig zu verstehen, über ein vielschichtiges kulturelles Höflichkeitssystem verfügen müssen, welches jedoch nicht einmal bei allen Angehörigen einer Sprachgemeinschaft gleich zu sein scheint. Des Weiteren unterliegen jegliche Ereignisse im fremdsprachigen Ausland, in denen reale Sprache produziert wird, zahlreichen Einflussfaktoren.⁷⁸¹

Er schlägt vor, derartige reale Materialien nach *Reichweiten* auszuwählen, wobei er sich an die Studien zum Basisvokabular des Englischen anlehnt. Dadurch soll der Blick für Idiosynkrasien und sehr spezifische Kontexte geschärft werden. Seiner Meinung nach müssten gleichzeitig entscheidende Spezifika des Handlungsumfelds thematisiert werden, weshalb er vorschlägt, solche Materialien im Unterricht zu vermitteln, die ein komplexes sprachliches Handlungsmuster mit einer realen, aber nicht zu einzelfallspezifischen sprachlichen Oberfläche realisieren. Dadurch soll verhindert werden, dass reale, aber nur unter besonderen Bedingungen eintreffende Handlungen in den ersten Lernjahren zum Modell werden. Schlickau schlägt vor, die neuen Medien, insbesondere die online zugänglichen Korpora, für die Auswahl der Materialien zu nutzen. In den Korpora könnte nach verschiedenen Mustern für Handlungsverbalisierungen gesucht werden, wie Schlickaus Beispiel der Schalterkommunikation in einer Postfiliale, und das häufigste Muster könnte den Schülern als übliches Modell vorgelegt werden. Aus diesen Belegen der Korpora wird außerdem ersichtlich, ob andere Handlungen, wie Begrüßung oder Verabschiedung auch zum üblichen Ablauf des Vorgangs gehören. Anhand deren Häufigkeit kann ermittelt werden, ob diese zusätzlichen Handlungen fakultativ oder obligatorisch sind. Hierdurch soll vermieden werden, dass wichtige Bestandteile solcher Vorgänge, wie beispielsweise

⁷⁸¹ Vgl. Schlickau (2009): op. cit., S. 36-40.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

Höflichkeitsfloskeln oder Grußformeln, in den Modellversionen für Schüler vergessen werden, da die ausgeführten Handlungen ansonsten von Muttersprachlern als inakzeptabel eingestuft werden könnten.⁷⁸²

Auch laut Aston ist die Nutzung von Korpora vor allem im kommunikativen Sprachunterricht von großer Bedeutung, da hier mit realen Materialien gearbeitet werden soll und nicht mit frei zum Zweck des Unterrichts erfundenen oder modifizierten Texten und Aufgaben. Es sollen reale Texte verwendet werden, um die natürlichen Sprachmuster nicht zu verzerren. Anhand der natürlichen Sprachmuster soll gezeigt werden, wie diese wirklich im Kontext gebraucht werden, um zu kommunizieren. Daher sind diese Texte auch motivierender für die Lerner.⁷⁸³

Im Sinne der engen Definition des kommunikativen Sprachunterrichts ist das Lernen eine Konsequenz des Aushandelns von Bedeutung in der Kommunikation. Die Hauptaufgabe des Unterrichts ist die Bereitstellung von Kontexten, in welchen der Lerner die Sprache verwenden kann, um zu kommunizieren, sei es durch die Produktion und Rezeption schriftlicher Texte oder durch mündliche Interaktion.⁷⁸⁴ Daher sind die Texte aus Korpora nicht nur Quellen für Beispielsätze in Form von Konkordanzen (siehe Kapitel 1.4.3.), sondern auch für Rede- oder Schreibanlässe. Auch ohne die Untersuchung von Texten nach linguistischen Gesichtspunkten, sollten diese realen Texte oder Textausschnitte gelesen werden, um darüber zu sprechen oder einen eigenen Text in der Fremdsprache zu produzieren.

López-Ferrero betont zum Ende ihrer Studie über den Gebrauch parataktischer Konjunktionen in einem Lernerkorpus im Vergleich zu demjenigen in einem Spezialkorpus die Bedeutung von Lernerkorpora für den Spracherwerb. Dabei weist sie ausdrücklich darauf hin, dass nicht nur im Zweit- oder Fremdspracherwerb, sondern auch für den Erwerb der Muttersprache die aus Lernerkorpora gewonnenen Erkenntnisse zur Entwicklung der kommunikativen Kompetenz beitragen können. Die Informationen aus einem Lernerkorpus sind von großem Nutzen, um Aspekte der Grammatik und des Sprachgebrauchs, die besonderer didaktischer Aufmerksamkeit bedürfen, zu erklären und können somit bei der Planung einer pädagogischen Grammatik helfen. Anhand eines

⁷⁸² Vgl. Schlickau (2009): op. cit., S. 41-44.

⁷⁸³ Vgl. Aston (2000): op. cit., S. 12.

⁷⁸⁴ Vgl. Aston (2000): op. cit., S. 14.

Korpus mit Texten von Schülern können deren Bedürfnisse ermittelt und damit herausgefunden werden, was ihnen zur Erlangung kommunikativer Kompetenz im akademischen Kontext fehlt.⁷⁸⁵ Auch Braun/Chambers geben an, dass Lernende im Sinne des kommunikativen Ansatzes Zugang zu realen Texten benötigen und nicht lediglich anhand konstruierter Beispiele lernen sollten, wie es in vielen Grammatiken und Lehrbüchern der Fall ist. Hierfür bieten sich ihrer Meinung nach Korpora als Datenquelle an.⁷⁸⁶

Die frühzeitige Arbeit mit realen Materialien leistet einen bedeutenden Beitrag zur Verringerung der Kluft zwischen der im Unterricht vermittelten Fremdsprache und der Erfahrung mit Muttersprachlern. Sofern die Schüler vor dem ersten Kontakt mit Muttersprachlern oder realen Medien in der Fremdsprache nur mit didaktisiertem Material gearbeitet haben, ist die Wahrscheinlichkeit der Enttäuschung oder Demotivation hoch. Die Konfrontation mit realen Materialien im Fremdsprachenunterricht fördert die kommunikative und interkulturelle Kompetenz durch den Umgang mit der Sprache und Kultur von Angehörigen der zielsprachlichen Umgebung.

Ciesielska-Ciupek spricht der Arbeit mit Korpora, bei der sich die Schüler mit realen Texten beschäftigen, nicht nur einen positiven Einfluss auf die kommunikative Kompetenz zu, sondern sie beschreibt ihre Schüler als autonomere und unabhängigere Lerner, nachdem sie längere Zeit selbstständig mit dem Internet und mit Korpora gearbeitet haben.⁷⁸⁷ In diesem Sinne wird auch die Lernerautonomie durch die Beschäftigung mit Korpora gefördert.

3.2.3. Der Begriff der Lernerautonomie

Die Lernerautonomie ist ein weiterer Aspekt des modernen Unterrichts. Die Schlagworte Eigenverantwortlichkeit und Selbstbestimmung kennzeichnen die Idealvorstellung des Fremdsprachenlernalers laut Grünewald/Küster bereits seit den reformpädagogischen Initiativen gegen Ende des 19. Jahrhundert, vor allem aber seit den 1990er Jahren.⁷⁸⁸

⁷⁸⁵ Vgl. López-Ferrero (2007): op. cit., S. 191-192.

⁷⁸⁶ Vgl. Braun/ Chambers (2009): op. cit., S. 332.

⁷⁸⁷ Vgl. Ciesielska-Ciupek (2003): op. cit., S. 530.

⁷⁸⁸ Vgl. Grünewald/ Küster (2009): op. cit., S. 226.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

Auch im GER (siehe Kapitel 3.1.2.) wird großer Wert auf selbstorganisiertes und selbstverantwortliches Lernen gelegt, da dies der Grundstein für das in der heutigen Gesellschaft so wichtige lebenslange Lernen ist. Lernstrategisches Wissen stellt dabei einen wichtigen Faktor für selbstverantwortliches Sprachenlernen dar, das heißt, die Lerner benötigen Lernstrategien und Lerntechniken, um sich individuelles Wissen über das Sprachenlernen anzueignen und es auch anwenden zu können.⁷⁸⁹

Van Lier kombiniert die Begriffe Autonomie und Authentizität im Hinblick auf den Fremdsprachenunterricht. Er sieht die Autonomie, die Authentizität und das Sprachbewusstsein als die drei Prinzipien eines idealen, modellhaften Lehrplans für Fremdsprachen und ist der Meinung, dass die Authentizität dabei sowohl das Ergebnis als auch der Ursprung des Sprachbewusstseins und der Autonomie ist. Schüler sollen im Unterricht dazu ermutigt werden, ihr Sprachbewusstsein zu entwickeln, selbstständig zu werden und in ihren Lernerfahrungen generell Authentizität anzustreben.⁷⁹⁰

Es stellt sich also die Frage danach, was unter Autonomie zu verstehen ist. Im *Wörterbuch für Erziehung und Unterricht* wird der Begriff der Autonomie wie folgt beschrieben:

„Autonomie bedeutet allgemein Eigengesetzlichkeit, Selbstständigkeit. Die autonome Erziehung betont die Unabhängigkeit [sic!] des Kindes von gesellschaftlichen Einflüssen und sein Heranwachsen ausschließlich auf Grund natürlicher Reifeprozesse und in ihm liegender Steuerungsmechanismen. Der Erzieher habe lediglich negative Einflüsse der Umwelt fernzuhalten. Eine andere Gewichtung erfährt Autonomie in der *Erziehung der Kinder und Jugendlichen zur Selbstständigkeit, insbesondere zu einem autonomen moralischen Urteilsvermögen* sowohl im Sinne eines Unterrichtsprinzips als auch speziell in Fächern wie Ethik und Religionslehre.“⁷⁹¹

Weiter wird angeführt, dass das Autonomieverständnis in der Auffassung Immanuel Kants von der freien Selbstbestimmung des Menschen gründet. Zum einen wird sie als selbstkritisch geleistete Unabhängigkeit von fremd bestimmenden und nicht hinterfragten Autoritäten und Traditionen gesehen, zum anderen als Unabhängigkeit von Neigungen und Begierden bei der Bestimmung moralisch relevanter Maximen und allgemeingültiger Gesetze. Somit wird die Autonomie zu einem der höchsten Lernziele

⁷⁸⁹ Vgl. Grünewald/ Küster (2009): op. cit., S. 84-86.

⁷⁹⁰ Vgl. Van Lier, Leo (1996): *Interaction in the Language Curriculum*. London: Longman, S. 10-20.

⁷⁹¹ Köck (2008): op. cit., S. 40.

auch im schulischen Umfeld, welches selten in vollem Umfang erreicht wird, aber als eines der übergeordneten Ziele den Unterricht bestimmt.⁷⁹²

In Bezug auf das Lernen von Sprachen kann der Begriff Autonomie laut Schmelter verstanden werden als „die Fähigkeit, die Verantwortung für sein eigenes Lernen zu übernehmen“. Der Lerner ist dann autonom, wenn er sich eigenständig ein Lernziel setzt, eine Art Handlungsplan (im Hinblick auf die Inhalte und Methoden) zur Erreichung desselben entwirft, diesen durchführt und schließlich evaluiert.⁷⁹³

Der Begriff *Lernerautonomie* an sich geht auf Holec zurück, welcher ihn im Hinblick auf die Erwachsenenbildung bereits im Jahr 1979 wie folgt definierte: „In the context with which we are dealing, the learning of languages, autonomy is consequently the *ability to take charge of one's own learning*.“ Lerner hätten seiner Meinung nach die Aufgabe, alle Entscheidungen bezüglich des Lernprozesses selbst zu treffen, vom Setzen eigener Lernziele über die Bestimmung von Inhalt und Progression, der Wahl geeigneter Techniken und Methoden bis hin zur Überwachung und Evaluierung des Lernfortschritts.⁷⁹⁴ Laut Schmenk sah Holec seine Definition der Lernerautonomie, im Sinne einer notwendigen Fähigkeit für selbstgesteuertes Lernen, jedoch nicht für den schulischen Fremdsprachenunterricht geeignet. Dennoch nahm eine verkürzte Version derselben ihren Umlauf in Kreisen des Fremdsprachenunterrichts. Hier wird Lernerautonomie daher, wie auch bei Schmelter, definiert als die Fähigkeit, das eigene Lernen selbst in die Hand zu nehmen beziehungsweise zu steuern. Dieser weit verbreiteten Beschreibung der Lernerautonomie liegt zwar Holec's Definition zugrunde, seine Vorstellungen werden jedoch häufig nicht berücksichtigt. Laut Schmenk gibt es derzeit allerdings verschiedenste Definitionen und Konzeptionen des Begriffs Lernerautonomie, weshalb man nicht von einer generell anzuwendenden Definition sprechen kann.⁷⁹⁵

⁷⁹² Vgl. Köck (2008): op. cit., S. 40.

⁷⁹³ Schmelter, Lars (2004): *Selbstgesteuertes und potenziell expansives Fremdsprachenlernen im Tandem*. Tübingen: Narr, S. 264.

⁷⁹⁴ Holec, Henry (1981): *Autonomy and Foreign Language Learning*. Oxford [u.a.]: Pergamon Press, S. 3.

⁷⁹⁵ Vgl. Schmenk, Barbara (2016): „Lernerautonomie und selbst gesteuertes Sprachenlernen“. In: Burwitz-Melzer, E./ Mehlhorn, G./ Riemer, C./ Bausch, K.-R./ Krumm, H.-J. (Hrsg.) (2016): *Handbuch Fremdsprachenunterricht*. Tübingen: A. Francke Verlag, S. 367-368.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

Grünewald/Küster sind der Meinung, für das Lernen in der Schule bedeute dies nicht, auf Lenkung und Vorgaben zu verzichten, da es in einem sozialen Kontext verankert ist und Lernende nicht von sich aus ihren eigenen Lernprozess autonom gestalten können. Die Schüler müssen erst nach und nach an die Autonomie herangeführt werden, wobei diese sich, ihrer Meinung nach, immer in einem dialektischen Spannungsverhältnis zwischen Autonomie und Heteronomie befinden.⁷⁹⁶

Laut Braun/Chambers unterstützt der Einsatz von Korpora die Entwicklung eben dieser Lernerautonomie, welche in konstruktivistischen und kognitiven Ansätzen gefordert wird. In den aktuellen Erkenntnissen der Sprachlernforschung werden vor allem Lernaktivitäten hervorgehoben, in denen die Schüler eine aktive Rolle einnehmen, wodurch ebenfalls die Eigenverantwortlichkeit der Lernenden für den Lernprozess gestärkt wird.⁷⁹⁷

Außerdem müssen in einem Sprachunterricht, der auf Lernerautonomie abzielt, die Rollen von Lehrer und Lerner neu formuliert werden. Trenchs Parera nennt den Lehrer in dieser Situation einen *asistente*. In diesem neuen Modell des Lernens ist der Lerner autonom und hat mehr oder weniger die Verantwortung für seinen eigenen Lernprozess. Als logische Konsequenz daraus verliert die Lehrkraft an Bedeutung und wird zu einem Assistenten, der dem Schüler dabei hilft, Materialien auszusuchen und zunehmend seine Fähigkeit, unabhängig zu sein, auszubauen. Ihrer Meinung nach ist die Autonomie dabei zum einen kognitiv, da jeder Lerner seinen eigenen Lernstil hat, zum anderen aber auch physisch, da er auch lernt, wenn die Lehrkraft nicht an seiner Seite steht. Zur Verwirklichung dieser Auffassung des Lernens dienen vor allem die neuen Medien und Technologien, welche das autonome Lernen überhaupt erst zulassen.⁷⁹⁸

Laut Rüschhoff/Wolff ist bei der Verwendung des Begriffs ‚Lernerautonomie‘ allerdings Vorsicht geboten. Im Zuge der Verwendung dieses Begriffs im Bereich des Lernens mit den Neuen Technologien verstehe man zunehmend einen Ersatz der Interaktion zwischen Lerner und Lehrer sowie derjenigen der Lerner untereinander durch die Lerner-Computer-Interaktion. Gemäß dieser Denkweise würde der Lerner mittels des

⁷⁹⁶ Vgl. Grünewald/ Küster (2009): op. cit., S. 226.

⁷⁹⁷ Vgl. Braun/ Chambers (2009): op. cit., S. 332.

⁷⁹⁸ Vgl. Trenchs Parera, Mireia (2001): „Tecnologías digitales para un cambio educativo en el aprendizaje de lenguas“. In: Trenchs Parera, M. (Hrsg.) (2001): *Nuevas Tecnologías para el Autoaprendizaje y la Didáctica de Lenguas*. Lleida: Milenio, S. 28-30.

Ersatzes des Lehrers durch den Computer und die uneingeschränkte Verfügbarkeit des Letzteren zu einem autonomen Lerner. Allerdings konnte für derartige Lernsituationen noch nicht einmal die Frage geklärt werden, wie offene Schülerfragen durch den Computer bewertet werden sollen.⁷⁹⁹

Rüschhoff/Wolff gehen von einem humanistisch geprägten Autonomiebegriff aus. Sie verstehen darunter ein allgemeines Erziehungsziel, das unter anderem in der Schule verwirklicht werden muss. Die Schüler müssen ein Gefühl für selbstständiges und eigenverantwortliches Lernen entwickeln.⁸⁰⁰ Auch Grünewald/Küster sehen die zunehmend selbstständige Aneignung der Sprache und das lebenslange Lernen als ein grundlegendes Ziel des Spanischunterrichts. Um den Lerner in die Lage zu versetzen, eigenständig lernen zu können, schlagen sie einige Mittel zur Förderung der Lernerautonomie vor. Unter anderem sollen förderliche Lernumgebungen geschaffen, Medien und Materialien für selbstständiges Arbeiten bereitgestellt, das Erkennen eigener Fähigkeiten vermittelt sowie *language awareness*, das heißt die Bewusstheit über Regularitäten des Sprachsystems, Kommunikationsstrategien und Sprache als Manipulationsmittel, gefördert werden. Außerdem wird die Entwicklung der Lernerautonomie durch handlungs- und kompetenzorientierten Unterricht unterstützt. Wenn Lerner selbst an der Planung und Gestaltung des Unterrichts beteiligt und dazu befähigt werden, Probleme zu lösen, sich eigenständig Wissen anzueignen und Entscheidungen zu treffen, werden sie gleichzeitig zu autonomen Lernern. Kleine Formen der Förderung der Schülerselbstständigkeit sind laut Grünewald/Küster beispielsweise die frühe Einführung des Umgangs mit Wörterbüchern, die Nutzung elektronischer Wortschatztrainer, die Selbstevaluation der Schüler mittels der Anwendung des GER, die Integration der Arbeit mit digitalen Plattformen sowie die Nutzung weiterer elektronischer Möglichkeiten auch zum außerschulischen Spanischlernen.⁸⁰¹ Die Korpora wären eine solche elektronische Möglichkeit, die die Arbeit mit Wörterbüchern erweitern kann und die im Fall der Korpora *Sacodeyl* (siehe Kapitel 2.4.3.1.) und *Backbone Spanish* (siehe Kapitel 2.4.3.2.) zusätzlich eine Art Lernplattform für Hörverstehensübungen darstellt, da hier nicht nur Texte und Audiodateien, sondern auch Aufgaben zu einzelnen Hörtexten verfügbar sind.

⁷⁹⁹ Vgl. Rüschhoff/ Wolff (1999): op. cit., S. 64.

⁸⁰⁰ Vgl. Rüschhoff/ Wolff (1999): op. cit., S. 65.

⁸⁰¹ Vgl. Grünewald/ Küster (2009): op. cit., S. 227-228.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

Im Zusammenhang mit der Lernerautonomie ist generell auch die Nutzung moderner Medien zu sehen. Jones/Stuhlmann/Zeyer geben an, dass junge Menschen aufgrund der Tatsache, dass sie außerhalb der Schule mit interaktiven Medien umgehen, aktiv am Lernprozess beteiligt werden möchten. Sie erwarten, dass sie im Unterricht nicht nur ‚Konsumenten‘, sondern auch ‚Produzenten‘ des Wissens sind, unabhängig davon, ob dieser mit oder ohne digitale Medien stattfindet. Diese Tendenz habe sich in den letzten Jahren noch intensiviert, eine Entwicklung, die damit zusammenhängt, dass annähernd jeder Jugendliche ein Mobiltelefon besitzt und die Mehrheit damit uneingeschränkten Zugriff auf das Internet hat.⁸⁰²

Von Grünewald/Küster wird zudem die emotionale Seite des Unterrichts angesprochen. Ein Unterricht, an dem die Schüler aktiv beteiligt werden und Lernerautonomie gefördert wird, führt zunehmend zu motivierteren Schülern und weckt in diesen positive Emotionen. Zusammen mit einer angstfreien Lernumgebung, der Schaffung von Erfolgserlebnissen, abwechslungsreichen Lernformen, der Neugier der Schüler und der Einbeziehung der spanischen Außenwelt können die Schüler so positiv beeinflusst werden.⁸⁰³

Des Weiteren trägt die Verwirklichung einer Binnendifferenzierung zur Förderung der Lernerautonomie bei. Schüler haben unterschiedliche kognitive, emotionale und motivationale Voraussetzungen, die sie in den Unterricht mitbringen, daher wird seit Jahren eine Binnendifferenzierung innerhalb der Klassen gefordert, die je nach den Lernvoraussetzungen individuell auf die Schüler beziehungsweise auf verschiedene Leistungsgruppen eingeht.⁸⁰⁴ Indem die Schüler bei der freien Arbeit mit Korpora selbst entscheiden, was sie genauer untersuchen beziehungsweise welche Texte auf welcher Kompetenzstufe sie lesen oder anhören, kann damit durchaus ein Beitrag zur Binnendifferenzierung geleistet werden.

⁸⁰² Vgl. Jones, Roger D./ Stuhlmann, Sebastian/ Zeyer, Tamara (2016): „Interaktives Fremdsprachenlernen: Potenziale und Herausforderungen“. In: Zeyer, T./ Stuhlmann, S./ Jones R. (Hrsg.) (2016): *Interaktivität beim Fremdsprachenlehren und –lernen mit digitalen Medien. Hit oder Hype?* Tübingen: Narr Francke Attempto, S. 12-16.

⁸⁰³ Vgl. Grünewald/ Küster (2009): op. cit., S. 229-230.

⁸⁰⁴ Vgl. Eisenmann, Maria (2016): „Binnendifferenzierung“. In: Burwitz-Melzer, E./ Mehlhorn, G./ Riemer, C./ Bausch, K.-R./ Krumm, H.-J. (Hrsg.) (2016): *Handbuch Fremdsprachenunterricht*. Tübingen: A. Francke Verlag, S. 358.

Ebenso betont Bernardini, die jahrelang mit italienischen Studenten der Übersetzungswissenschaften englische Korpora zur eigenständigen Arbeit nutzte, die Bedeutung der Autonomie des Lernalers im Lernprozess. Ihrer Meinung nach sollen die Lerner Verantwortung für ihren eigenen Lernprozess übernehmen. Sie bevorzugt dabei Aufgaben im Sinne des *discovery learning*, wie in Kapitel 3.3.1.2. zu sehen sein wird, da diese so konzipiert sind, dass sie lernerzentriertes, an die jeweilige Person angepasstes Lernen mit offenem Ausgang begünstigen und dadurch die Lernerautonomie fördern.⁸⁰⁵ Die didaktische Vorgehensweise im Umgang mit Studenten ist eine andere als mit Schülern, das soll hier ganz klar herausgestellt werden, doch im Fremdsprachenunterricht an der Schule sollte das entdeckende Lernen ebenso gefördert werden wie an der Hochschule. Auch John Osborne gibt in seinem Aufsatz „Integrating corpora into a language-learning syllabus“ neben der Authentizität die Lernerautonomie als Hauptgrund für die Verwendung korpusbasierter Ansätze des Fremdsprachenunterrichts an.⁸⁰⁶

Alonso Pérez-Ávila ist ebenfalls der Ansicht, dass die Anwendung von Korpora im Fremdsprachenunterricht wesentliche Prinzipien kommunikativer Ansätze unterstützt, unter anderem die Autonomie des Lernalers bezüglich der Lehrkraft, eine Variation im Lernprozess und die Erhöhung der Motivation aufgrund der Unvorhersehbarkeit des Ergebnisses bei sogenannten *discovery activities*.⁸⁰⁷

Wenn die Schüler selbstständig in einem Korpus nach Lexemen, Kollokationen oder Synonymen suchen, können sie selbst bestimmen, wie lange sie die Suchergebnisse eines beliebigen Lexems analysieren, wie viele und welche Suchanfragen sie formulieren sowie welche Ergebnisse ihnen wichtig erscheinen. Sofern der Lehrer die Schüler nicht gängelt oder ihnen Aufgaben und Suchausdrücke genau vorgibt, kann sich der einzelne Schüler frei im Korpus bewegen und sich von einem Ergebnis oder einem beobachteten Sachverhalt zum nächsten führen lassen. Der Lehrer sollte dabei eine angemessene Dauer der Übung wählen, damit die Schüler auch Suchausdrücke ausprobieren können, die nicht unmittelbar zum späteren Ergebnis beitragen, sondern einfach nur aus Interesse

⁸⁰⁵ Vgl. Bernardini, Silvia (2004): „Corpora in the classroom. An overview and some reflections on future developments“. In: Sinclair, J. McH. (2004): *How to use corpora in language teaching*. Amsterdam/Philadelphia: Benjamins, S. 27.

⁸⁰⁶ Vgl. Osborne, John (2003): „Integrating corpora into a language-learning syllabus“. In: Lewandowska-Tomaszczyk, B. (Hrsg.) (2003): *PALC 2001: Practical Applications in Language Corpora*. Frankfurt am Main: Lang, S. 478.

⁸⁰⁷ Vgl. Alonso Pérez-Ávila (2007): op. cit. (06.06.2017), S. 6.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

eingetragen werden. Das Interesse der Schüler stellt einen der wichtigsten Faktoren für effektives und nachhaltiges Lernen dar. Am Ende der Unterrichtseinheit sollten die Ergebnisse miteinander verglichen und besprochen sowie wichtige Sachverhalte eventuell festgehalten werden. Im Fall einer freien Besprechung der Ergebnisse jedes Schülers hat niemand die Verpflichtung sich während seiner Suche genau an einen Zeitplan oder an das erwünschte Ergebnis halten zu müssen, sondern kann länger bei interessanten Beispielen verweilen. Es ist jedoch klar, dass eine derart freie Unterrichtsgestaltung nicht immer funktioniert. Die Effektivität der Korpusarbeit ist abhängig von der jeweiligen Klasse, der Tageszeit und dem Alter der Schüler. Allerdings gibt es auch bei anderen Übungsformen keine Garantie dafür, dass die Unterrichtsstunde genauso verläuft, wie der Lehrer sie geplant hat, da der Ausgang der Stunde nicht vorhergesehen werden kann.

In diesem Sinn kann der Einsatz von Korpora zur selbstständigen Suche der Schüler in fremdsprachlichen Texten durchaus zur Förderung der Autonomie beitragen. Wenn der Schüler selbstständig im Korpus forscht und anhand von Konkordanzen Regelmäßigkeiten erkennt sowie infolgedessen Regeln ableitet, wird er zunehmend ein autonomer Lerner, der aktiv an seinem eigenen Lernprozess beteiligt ist. Die Nutzung von Korpora muss jedoch zunächst erklärt und eingeübt werden, indem der Lehrer den Schülern das Wissen an die Hand gibt, das diese benötigen, um sich schließlich selbstständig damit beschäftigen zu können. So entwickeln die Schüler allmählich das von Rüschoff/Wolff angeführte Gefühl für selbstständiges und eigenverantwortliches Lernen.

Nachdem diese drei für den Fremdsprachenunterricht bedeutenden Kriterien, welche durch die Korpusarbeit gefördert werden können, geklärt sind, stellt sich nun aber die Frage danach, wie die Korpora letztendlich in den Unterricht integriert werden können. Die verschiedenen Nutzungsmöglichkeiten sollen im folgenden Unterkapitel besprochen werden.

3.3. Einsatzmöglichkeiten von Korpora im Fremdsprachenunterricht

Leech unterscheidet im einleitenden Kapitel von *Teaching and Language Corpora* zwischen der direkten und indirekten Nutzung von Korpora für den Sprachunterricht sowie der Entwicklung von unterrichtsorientierten Korpora. In den letzten Bereich fällt

die Entwicklung von Korpora für den Unterricht von LSP (*Languages for Specific Purposes*) sowie von Korpora für den Erst- und Zweitspracherwerb. Unter dem indirekten Gebrauch versteht Leech das Verlegen von Nachschlagewerken, die Entwicklung von Unterrichtsmaterialien sowie die Erstellung von Sprachtests. Die direkten Verwendungsweisen hingegen sind *Teaching about*, *Exploiting to teach* und *Teaching to exploit*.⁸⁰⁸

3.3.1. Direkte Nutzung von Korpora

Fligelstone unterscheidet die eben genannten drei Bereiche der direkten Verwendung von Korpora wie folgt: Der Bereich *Teaching about* beschäftigt sich mit dem Unterricht über Korpora und Korpuslinguistik, also mit deren Geschichte, Theorie, Zielen und Einwänden. Dieser Zweig bezieht sich hauptsächlich auf die Lehre an Universitäten, wo die Korpuslinguistik zunehmend in die Curricula der universitären Linguistik aufgenommen wird.⁸⁰⁹ Die Korpuslinguistik wird hier als eine Subdisziplin neben Syntax und Pragmatik unterrichtet.⁸¹⁰ In den vorherigen Kapiteln über Korpora und Korpuslinguistik sowie über die Kategorisierung der Korpora wurden bereits eben diese Gebiete ausführlich besprochen. Im Sinne des bei *Teaching about* gemeinten Unterrichts über Korpuslinguistik und Korpora würden diese Informationen an die Schüler beziehungsweise Studenten weitergegeben. Da es sich hierbei jedoch lediglich um die Theorie der Korpuslinguistik handelt, sollte dieser Teil der direkten Beschäftigung mit Korpora im Unterricht eher kurz ausfallen. Lediglich in einem speziellen Seminar in der Oberstufe könnte genauer auf die Theorie eingegangen werden.

Um jedoch Korpora im und für den Unterricht nutzen zu können, müssen die Lehrer mit dem korpusbasierten Ansatz vertraut sein. Im englischsprachigen Bereich gibt es beispielsweise eine Studie von Allan zur Förderung des Sprachbewusstseins von

⁸⁰⁸ Vgl. Leech, Geoffrey (1997): „Teaching and Language Corpora: a Convergence.“ In: Wichmann, A./ Fligelstone, S./ McEnery, T./ Knowles, G. (Hrsg.) (1997): *Teaching and Language Corpora*. London: Longman, S. 5-6.

⁸⁰⁹ Vgl. Fligelstone, Steven (1993): „Some reflections on the question of teaching, from a corpus linguistics perspective“. In: *ICAME Journal. Computers in English Linguistics*. No. 17. April 1993. Bergen: ICAME, S. 98-102. Online Publikation: http://clu.uni.no/icame/archives/No_17_ICAME_Journal_index.pdf (08.06.2017).

⁸¹⁰ Vgl. McEnery/ Xiao (2010): op. cit., S. 370.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

Engischlehrern an weiterführenden Schulen in Hongkong.⁸¹¹ Eine andere Studie von Conrad beschäftigt sich mit *linking adverbials* (beispielsweise *therefore* oder *in other words*) und lässt die Autorin darauf schließen, dass Lehrer die Korpora nicht nur für die Arbeit mit Konkordanzen oder für lexikalische beziehungsweise lexikogrammatikalische Analysen im Klassenzimmer verwenden sollten, um die volle Bandbreite des korpusbasierten Ansatzes zu nutzen.⁸¹² Conrad sowie O’Keeffe/Farr⁸¹³ sind der Meinung, die Korpuslinguistik sollte in die Ausbildung von Sprachlehrern aufgenommen werden, um die Forschungskompetenzen und das Sprachbewusstsein derselben weiterzuentwickeln.

Exploiting to teach, also die Nutzung der Korpora für den Unterricht, ist laut Fligelstone der am weitesten entwickelte, aber gleichzeitig am wenigsten erschöpfend ausgebaute Bereich, was für ihn keinen Widerspruch darstellt, da dieser Bereich seiner Meinung nach unerschöpflich ist.⁸¹⁴ Für McEnery/Xiao bedeutet *Exploiting to teach*, dass ein korpusbasierter Ansatz für Sprach- und Sprachwissenschaftskurse (beispielsweise der Soziolinguistik oder Diskurslinguistik) angewandt wird, welcher ansonsten ohne Korpora unterrichtet werden würde.⁸¹⁵

Der dritte Bereich der direkten Verwendungsweisen, *Teaching to exploit*, beschäftigt sich mit dem Unterricht über die Nutzung von Korpora, das heißt mit der Lehre von deren Nutzungsmöglichkeiten. Es handelt sich hierbei nach Fligelstones Meinung um den wichtigsten der drei Bereiche, mit dem man Schüler oder Studenten nachhaltig in ihrem Lernverhalten unterstützen kann, sodass diese die Korpora auch für ihren eigenen, individuellen Lernprozess nutzen.⁸¹⁶ Hier wird den Lernern praktisches Know-how an die Hand gegeben, sodass sie Korpora für ihre eigenen Zwecke nutzen können.⁸¹⁷

Aufgrund der Tatsache, dass die Inhalte des Bereichs *Teaching about* bereits in den ersten beiden Kapiteln beschrieben wurden, werden im Folgenden lediglich die Bereiche *Exploiting to teach* und *Teaching to exploit* genauer ausgeführt werden.

⁸¹¹ Vgl. Allan, Quentin Grant (1999): „Enhancing the language awareness of Hong Kong teachers through corpus data“. In: *Journal of Technology and Teacher Education*, 7 (1), S. 57–74.

⁸¹² Vgl. Conrad, Susan (1999): „The importance of corpus-based research for language teachers“. In: *System*, 27, S. 1–18.

⁸¹³ Vgl. O’Keeffe, Anne/ Farr, Fiona (2003): „Using language corpora in initial teacher education: pedagogic issues and practical applications“. In: *TESOL Quarterly*, 37 (3), S. 389–418.

⁸¹⁴ Vgl. Fligelstone (1993): op. cit., S. 101.

⁸¹⁵ Vgl. McEnery/ Xiao (2010): op. cit., S. 370.

⁸¹⁶ Vgl. Fligelstone (1993): op. cit., S. 100–101.

⁸¹⁷ Vgl. McEnery/ Xiao (2010): op. cit., S. 370.

3.3.1.1. *Exploiting to teach*

Der Bereich *Exploiting to teach* beschäftigt sich also mit der Nutzung der Korpora für den Unterricht, wobei sich, wie auch im traditionellen Unterricht, zwei Herangehensweisen anbieten. Es wird zwischen der induktiven und der deduktiven Methode unterschieden. Im *Wörterbuch für Erziehung und Unterricht* wird die induktive Unterrichtsmethode wie folgt beschrieben:

„Durch die induktive Unterrichtsmethode wird die Wahrnehmungswelt des Schülers verarbeitet. Sie ist ein Verfahren, das Einzelerfahrungen durch weitere Elemente, durch negative und positive Beispiele vermehrt, um eine Grundeinsicht zu vermitteln. Besonders bei Kindern und Jugendlichen ist das induktive Verfahren angebracht, da es von Erfahrung und dem jeweiligen Einzelfall ausgeht und zur Regel, Definition oder Gesetzmäßigkeiten führt. In der Übungsphase wird die Induktion wiederum durch die Deduktion ergänzt, wenn der Beweis für die Richtigkeit der gefundenen Regel bzw. Gesetzmäßigkeit durch Übungsaufgaben angetreten wird.“⁸¹⁸

Im Gegensatz dazu führt die deduktive Methode von der Regel zum Beispiel, der Schüler wird also vom Allgemeinen zum Besonderen geleitet.⁸¹⁹ Siepmann rät, mittels deduktiver Aufgaben in die Korpusarbeit einzusteigen und den Schülern erst später, wenn sie Erfahrung mit dem Korpus gesammelt haben, induktive Aufgabenstellungen vorzulegen, bei denen das Datenmaterial unmittelbar gesichtet wird, um Regularitäten zu ermitteln. Hierbei sollte die Lehrkraft zunächst noch Hilfestellung leisten, indem sie beispielsweise die Daten in eine für die Lösung der Aufgabe förderliche Ordnung bringt. Zudem rät er, die Schüler nicht zu überfordern, denn obgleich die Analyse von Konkordanzen schnell zur Routine würde, sei sie doch eine probate Aufgabenstellung. Er rät zudem von Aufgaben ab, welche eine Analyse unterschiedlicher Bedeutungen von hochfrequenten polysemen Lexemen wie dem englischen *take* verlangen. Außerdem sollte sowohl bei induktiven als auch bei deduktiven Aufgabenstellungen größter Wert darauf gelegt werden, dass sich die Schüler der Kontextgebundenheit sprachlicher Äußerungen bewusst sind.⁸²⁰

Im Unterricht kann der Lehrer demnach auf zweierlei Arten mit Konkordanzen umgehen. Zum einen kann er im Sinne des deduktiven Lernens zunächst allgemeine Regeln bezüglich eines grammatischen Sachverhalts erläutern und den Schülern, nach dieser allgemeinen Einführung in das Thema, Konkordanzen aus einem

⁸¹⁸ Köck (2008): op. cit., S. 212.

⁸¹⁹ Vgl. Köck (2008): op. cit., S. 82.

⁸²⁰ Vgl. Siepmann (2009): op. cit., S. 327.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

fremdsprachlichen Korpus zur Verfügung stellen. Bei der Arbeit mit den Konkordanzen können diese entweder nur als Beispiele zur Veranschaulichung oder aber als Grundlage für eine didaktische Übung dienen, bei welcher der Lehrer die Konkordanzen zuvor in der Art verändert, dass die Schüler Lücken füllen oder Fehler erkennen müssen. Zum anderen kann der Lehrer, dem induktiven Lernen gemäß, den Schülern zuerst Konkordanzen zur Verfügung stellen und erst im zweiten Schritt werden aus diesem fremdsprachlichen Material Regeln und Leitsätze entwickelt, welche die Schüler dann auch auf andere Beispiele übertragen.

Osborne kombiniert deduktives und induktives Lernen, indem er den Schülern zunächst, der deduktiven Methode folgend, generelle Regeln an die Hand gibt, deren Gültigkeit die Schüler mithilfe von Konkordanzen überprüfen sollen. Nach der Durchsicht der Konkordanzen sollen die Schüler dann für diejenigen Beispiele, auf welche die Regeln nicht zutreffen, im Sinne der induktiven Methode eigene Merksätze erstellen.⁸²¹ Durch diese Kombination der beiden Methoden sind die Schüler auf unterschiedliche Weise gefordert, und sie sind motiviert, da sie selbstständig arbeiten.

Konkordanzen können im Unterricht ebenso gut dem Einstieg in ein neues Lerngebiet dienen. Der Lehrer kann die Schüler zunächst fragen, was sie als die typische sprachliche Umgebung eines bestimmten Lexems erachten, und sie anhand der Konkordanzen Hypothesen aufstellen lassen. Später können die Schüler ihre Ergebnisse auch anhand einer Schulgrammatik überprüfen und bei Unklarheiten mit dem Lehrer Rücksprache halten. Nachdem die Schüler die sprachlichen Regelmäßigkeiten erkannt, einander zugeordnet und generalisiert haben, können sie sich selbst einige Beispiele ausdenken oder Aufgaben des Lehrers lösen und schließen so an den induktiven Lernprozess eine deduktive Übungsphase an. Induktives und deduktives Lernen können somit im Hinblick auf die Arbeit mit Korpusdaten durchaus gewinnbringend miteinander kombiniert werden.

Wofür die Konkordanzen auch immer verwendet werden, so handelt es sich dabei auf jeden Fall um reale Sätze, die Erklärungen zu Lexik, Grammatik oder auch zu einem kulturellen Thema ergänzen, und keine ad hoc vom Lehrer ausgedachten Beispiele,

⁸²¹ Vgl. Osborne, John (2000): „What can students learn from a corpus?: building bridges between data and explanation“. In: Burnard, L./ McEnery, T. (Hrsg.) (2000): *Rethinking language pedagogy from a corpus perspective: papers from the Third International Conference on Teaching and Language Corpora*. Frankfurt am Main: Peter Lang, S. 165-172.

welche bisweilen künstlich klingen können und nicht der Art und Weise entsprechen, wie Muttersprachler sich ausdrücken würden. Zudem schränkt der Lehrer das Input der Schüler stark ein, wenn er das einzige sprachliche Modell im Unterricht bleibt. Er beeinflusst dadurch die Qualität des Inputs und damit zwangsläufig auch die sprachlichen Fähigkeiten der Schüler. Die Texte der Korpora stammen aus unterschiedlichsten Quellen, da beispielsweise ein Referenzkorpus für gewöhnlich mündliche und schriftliche Texte formeller sowie informeller Sprache enthält. Des Weiteren ermöglicht das Lernen anhand von Konkordanzen aus einem Korpus das Erlernen von typischen Kontexten, Einschränkungen in der Verwendung und pragmatischen Charakteristika von Lexemen.⁸²² Es ist immer von Vorteil, die gelernten Grundregeln anhand realer Daten zu verifizieren, um den Schülern die reale Sprachverwendung von Muttersprachlern vor Augen zu führen und sie somit zum Lernen zu motivieren.

Wie bereits erwähnt, enthalten vor allem Referenzkorpora Texte verschiedenster Herkunft. Anhand dieser Korpora oder auch verschiedener Spezialkorpora können bestimmte Register und Sprachvarianten miteinander verglichen werden. So können Unterschiede zwischen wissenschaftlichen und mündlichen Texten oder Spezialwortschatz aus unterschiedlichen Genres oder Disziplinen ermittelt werden.⁸²³ Des Weiteren ist es möglich, Korpora zu stilistischen und thematischen Untersuchungen heranzuziehen. So kann ein Korpus auch der Veranschaulichung von Stilmitteln dienen, beispielsweise bei der Behandlung eines literarischen Werks, oder der Suche nach Lexemen, welche zu einem Thema gehören. Ein Beispiel hierfür ist das Thema ‚Liebe‘ bei McCullough, der mittels eines Korpus verschiedene Werke nach *love*, dessen Synonymen und Antonymen durchsucht und so herausgefunden hat, wo er Informationen zu diesem Thema finden kann.⁸²⁴

Ein Vorteil beim Vergleich von Texten mithilfe von Korpora ist die Möglichkeit, problemlos die *type-token*-Relation (siehe Kapitel 1.4.1.) zu ermitteln und dadurch Aussagen über die Größe des verwendeten Wortschatzes zu treffen. Die durchschnittliche Wortlänge sowie häufige Kollokationen können erörtert und anschließend kann darüber

⁸²² Vgl. McCullough (2001): op. cit., S. 129.

⁸²³ Vgl. Tolchinsky, Liliana (2014): „El uso de corpus lingüísticos como herramienta pedagógica“. In: *Textos de Didáctica de la Lengua y de la Literatura*, enero 2014, núm. 65, S. 14-15. Online Publikation: <http://diposit.ub.edu/dspace/bitstream/2445/66774/1/653407.pdf> (05.06.2017).

⁸²⁴ Vgl. McCullough (2001): op. cit., S. 131.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

diskutiert werden.⁸²⁵ Anhand der Arbeit mit einem Korpus ist das Konzept der Kollokation (siehe Kapitel 4.1.1.) für Schüler besser greifbar, da sich ihnen im Korpus zahlreiche Kontexte präsentieren. Somit ist es möglich, dass sie die Kollokationen besser verstehen und sie sehen, wie wichtig es ist, Lexeme im Kontext zu lernen.⁸²⁶

Eine Möglichkeit der Einbettung von Konkordanzen in den Unterricht ist die Kombination allgemeiner Lehrbuchgrundsätze mit Korpusmaterial. Der Lehrer führt im Unterricht beispielsweise gewisse grammatikalische Regeln ein und lässt die Schüler dann im Korpus oder in vorher ausgewählten Konkordanzen nach Beispielen und Beweisen für diese Regeln suchen. Die Schüler sollen hier anhand der Belege die zuvor gelernten Regeln entweder bestätigen oder entkräften.

Eine andere Verwendungsweise ist die Suche nach Regeln anhand der Daten aus dem Korpus. So stellt der Lehrer mithilfe allgemeiner, in Schulbüchern vermittelter Grammatikregeln Behauptungen auf und lässt die Schüler diese Behauptungen anhand eines Korpus oder anhand einiger Konkordanzen überprüfen. Nach dieser Phase der Eigenaktivität seitens der Schüler, werden die Hypothesen noch einmal gemeinsam auf deren Richtigkeit hin geprüft und die gültigen Regeln werden festgehalten. So kann die Lehrkraft zum Beispiel bei der Behandlung der Auslöser des *subjuntivo* die Adverbien des Zweifels angeben (*acaso, tal vez, quizá(s), a lo mejor*) und den Schülern lediglich mitteilen, dass manche mit Indikativ und manche mit Indikativ oder *subjuntivo* verwendet werden, je nachdem wie sicher sich der Sprecher ist.⁸²⁷ Die Schüler können dann in einem beliebigen Korpus (beispielsweise in einem Referenzkorpus wie CREA, CORPES oder CdE; siehe Kapitel 2.1.1.) nach diesen Adverbien suchen und ermitteln, welche davon den *subjuntivo* auslösen. Bei dieser Art der Suche muss lediglich das jeweilige Adverb in die Eingabezeile der Suchmaske geschrieben werden (ohne weitere Einstellungen oder Angaben), wodurch diese Suchmöglichkeit auch bei Schülern angewandt werden kann, die noch wenig Erfahrung mit Korpora haben. Je nach Vorgehensweise im Unterricht kann die Lehrkraft auch die vier Adverbien nennen und nur angeben, dass die Schüler herausfinden sollen, welche den *subjuntivo* und welche den Indikativ nach sich ziehen,

⁸²⁵ Vgl. Tolchinsky (2014): op. cit., S. 14-15.

⁸²⁶ Vgl. McCullough (2001): op. cit., S. 129.

⁸²⁷ Vgl. Rathsam, Kathrin/ Schleyer, Jochen (2014): *Spanische Grammatik für die Mittel- und Oberstufe*. Berlin: Cornelsen, S. 95.

oder sie kann richtige beziehungsweise falsche Behauptungen aufstellen, welche überprüft werden sollen. So könnten die Hypothesen folgende sein:

- Nach *quizá(s)* steht immer der *subjuntivo*. (Falsch: Indikativ oder *subjuntivo*)
- Nach *acaso* steht immer der Indikativ. (Richtig)
- *A lo mejor* kann von Indikativ oder *subjuntivo* gefolgt werden. (Falsch: Indikativ)
- *Tal vez* steht immer mit dem *subjuntivo*. (Falsch: Indikativ oder *subjuntivo*)

Sollten im Unterricht nicht genügend Computer zur Verfügung stehen oder die Schüler noch nicht genug Erfahrung mit Korpora gesammelt haben, kann der Lehrer den Schülern auch zuvor aus einem Korpus extrahierte Konkordanzen auf einem Arbeitsblatt zur Verfügung stellen, anhand dessen die Schüler die Regeln erarbeiten sollen. Sofern mit vorher von der Lehrkraft ausgesuchten Konkordanzen gearbeitet wird, können die festgehaltenen Regeln immer Bezug auf bestimmte Konkordanzen nehmen, um den Schülern die Möglichkeit zu geben, die Grundsätze jederzeit auch anhand eines oder mehrerer Beispiele zu wiederholen. Ein Vergleich bestimmter Beispiele, um diese zu differenzieren und Merksätze zur Verwendung dieser Grammatikphänomene aufzustellen, ist ebenso gut möglich, wie in Kapitel 4.2.5. anhand von Konkordanzen mit verschiedenen Auslösern des *subjuntivo* zu sehen sein wird.

McCullough gibt zudem an, dass anhand eines Konkordanz- und eines Textverarbeitungsprogramms in kürzester Zeit Arbeitsmaterial zu einem aktuellen Text, beispielsweise einem Zeitungsartikel, erstellt werden kann. Zunächst wird die digitale Textdatei in das Konkordanzprogramm geladen und eine alphabetische Liste aller darin enthaltenen Lexeme erstellt. Anhand dieser Liste kann relevantes Vokabular zum jeweiligen Thema (bei McCullough aus dem juristischen Bereich) ausgewählt werden, das mittels des Konkordanzprogramms gesucht und als Konkordanzliste ausgegeben wird. Infrage kommende Sätze werden daraufhin in ein Textverarbeitungsprogramm kopiert und dort eine Lückenübung erstellt, bei der die Schüler das fehlende Vokabular einsetzen sollen. Ebenso kann nach Relativpronomen (in McCulloughs Beispiel die englischen Relativpronomen *which*, *who*, *whom* und *whose*) gesucht und eine passende Übung dazu entwickelt werden. Das Konkordanzprogramm erspart der Lehrkraft dabei das mehrmalige Lesen des Texts zur Suche nach Lexemen und Beispielsätzen. Außerdem erleichtert es die Erstellung von Lückenübungen.⁸²⁸ Beispiele für die kombinierte

⁸²⁸ Vgl. McCullough (2001): op. cit., S. 132-133.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

Nutzung eines Konkordanz- und eines Textverarbeitungsprogramms finden sich in Kapitel 4.7., einem Exkurs zur Erstellung eines eigenen Korpus.

Korpora können die Lehrkraft auch bei der Erstellung von Schulaufgaben und Stegreifaufgaben unterstützen. Sie liefern nicht nur reales Sprachmaterial, sondern bieten aufgrund ihrer vorteilhaften Gestaltungsweise viele Möglichkeiten zur einfachen Entwicklung von Aufgaben. Zunächst können Korpora als Datenbank für reale spanische Texte dienen, welche der Lehrer für den Unterricht oder auch für Schulaufgaben verwenden kann. Durch die Möglichkeit der Suche nach Schlagwörtern über das Eingabefeld kann nach Texten aus bestimmten Themengebieten gesucht und aus den Ergebnissen ein passender ausgewählt werden. Bei der Nutzung von Korpora zur Suche von Texten oder Textausschnitten ist es vorteilhaft, dass bei den meisten Korpora nach Genre oder geografischer Herkunft sowie eventuell nach Geschlecht und Alter der Sprecher gesucht werden kann.

Des Weiteren können Beispielsätze und Sätze für Übungen aus den Konkordanzen des Korpus entnommen werden. Hierbei bietet das Korpus eine relativ zeitsparende Methode der Erstellung von Lückentexten oder Vervollständigungsübungen, indem das jeweilige Lexem gesucht, ein passender Beispielsatz in die Übung integriert und das betreffende Lexem wieder gelöscht wird.

Auch bei der Korrektur von Schulaufgaben und Stegreifaufgaben sind die Korpora für Lehrer nützlich. Häufig ist sich eine nicht-muttersprachliche Lehrkraft nicht sicher, ob ein bestimmter Ausdruck oder eine grammatische Struktur wirklich falsch oder ihr selbst nur nicht vertraut ist. Oft gestaltet sich die Suche in Grammatiken und Wörterbüchern als sehr schwierig und zeitraubend, besonders wenn es sich um verzwickte Konstruktionen handelt, bei denen die Lehrkraft nicht auf Anhieb weiß, in welchem Kapitel des Nachschlagewerks die Lösung zu finden ist. In diesem Fall kann ohne großen Zeitaufwand im Korpus gesucht und je nach dem Ergebnis ein Urteil gefällt werden. Allerdings muss der Lehrer dabei beachten, dass nicht jede Struktur, die im Korpus nicht vorkommt, auch nicht existiert. Ein Korpus ist immer eine begrenzte Sammlung von Texten, daher kann für seltene Strukturen unter Umständen kein Beleg vorhanden sein. Allerdings finden sich für die meisten Konstruktionen dennoch Beispiele. Die Arbeit mit Korpora ist bei der Korrektur von Texten vor allem dann sehr hilfreich, wenn sich der Lehrer bei einer Formulierung unsicher ist, die Ergebnisse der

Suche aber eindeutig für die Existenz der vom Schüler verwendeten sprachlichen Struktur sprechen.

Wenn die Lehrkraft andererseits verunsichert ist, weil der Großteil der Schüler in den Aufsätzen schreibt **Trabaja como un vendedor*, so kann sie innerhalb kürzester Zeit im CREA nachsehen (vorzugsweise in der älteren, nicht-annotierten Version, siehe Kapitel 2.1.1.1.) und hier sowohl den obigen Satz als auch das korrekte Pendant *Trabaja como vendedor* eingeben. Sie wird einige Sätze für die zweite Version erhalten, jedoch keine Treffer für den ersten Satz. Andererseits ist man sich als Lehrer seiner Sache manchmal sehr sicher und hält bestimmte Konstruktionen der Schüler für falsch, wobei sich bei manchen ein Blick in ein Korpus lohnen würde. So geht man beispielsweise davon aus, dass der deutsche Ausdruck ‚in den vergangenen Tagen‘ im Spanischen *en los días pasados* heißt, einige Schüler schreiben in ihren Texten jedoch *en los pasados días*. Die nicht-annotierte Version des CREA führt hier fast ebenso viele Treffer für die zweite Version auf wie für die erste. Zur Ermittlung der Korrektheit dieses Ausdrucks bräuchte die Lehrkraft zwar nicht unbedingt ein Korpus, doch dieses liefert sogleich ein Ergebnis in Zahlen und lässt Vergleiche verschiedener Verwendungsweisen zu. Zudem können Korpora sehr aufschlussreich sein bei der Nutzung bestimmter Lexeme und Ausdrücke, welche in Wörterbüchern nur einzeln, ohne Kontext aufgeführt sind. So wird bei der Suche in Korpora wie CREA oder CORPES (siehe Kapitel 2.1.1.2.) klar, dass die Ländernamen *Argentina* und *Perú* sowohl mit als auch ohne bestimmten Artikel verwendet werden können. Wird jedoch *Colombia* mit dem bestimmten Artikel *la* kombiniert, dann nur, wenn hinter dem Substantiv eine Einschränkung desselben steht, wie *en la Colombia actual*. Ansonsten steht *Colombia* immer ohne Artikel. Auch die korrekte Nutzung von *etcétera* am Ende von Aufzählungen ist allein bei einem Blick in ein Wörterbuch nicht klar. Da in diesem nur der Ausdruck allein aufgeführt wird, ist nicht ersichtlich, dass das spanische *etcétera* anders als das deutsche mit Komma an die vorhergehende Aufzählung angehängt wird. Sofern der Ausdruck *etcétera* jedoch in ein Korpus wie das CORPES eingegeben wird, ist anhand der Konkordanzen schnell zu erkennen, dass man im Spanischen eben dieses Komma benötigt.

Korpora können von Lehrern auch sehr gut zur Korrektur von Vokabellückentexten verwendet werden. Wenn unklar ist, ob ein Lexem überhaupt existiert, kann es im Korpus nachgeschlagen werden und ist dort eventuell sogar aufzufinden, obwohl es in

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

Wörterbüchern noch nicht vermerkt ist. Auf diese Weise kann die unnötige Korrektur eigentlich korrekter Lexeme verhindert werden. Ein Beispiel ist ein Vokabeltest aus einer 10. Klasse (Spanisch als 3. Fremdsprache, 3. Lernjahr), bei welchem ein Schüler als Gegenteil von *casarse* das Lexem *descasarse* einsetzte. Zunächst würde dieses Lexem wohl als falsch markiert werden, da es sich in den gängigen Wörterbüchern nicht auffinden lässt. Eine Suche im CREA ergibt jedoch 2 Treffer für *descasarse*, 3 Treffer für *descasar* und weitere 2 Treffer für *descasa*. Im CdE:Orig (siehe Kapitel 2.1.1.3.) finden sich 3 Treffer für *descasar*. Verhütungsmittel sind allerdings *anticonceptivos* und die im Internet zu findende Form *anticonceptiva* anstatt des Ausdrucks *pildora anticonceptiva* ist in den Korpora CREA und CdE:Orig nicht zu finden. Es ist also fraglich, ob die Lehrkraft auf diese Form wirklich die volle Punktzahl geben sollte. Die Verwendung eines Korpus kann damit die klassische Korrektur mit Grammatiken und Wörterbüchern unterstützen oder zum Teil sogar ersetzen und dadurch die Korrekturzeit verkürzen.

Allerdings können nicht nur Lehrkräfte Korpora für die Korrektur verwenden, sondern auch die Schüler selbst oder zumindest mit der Lehrkraft zusammen. Anhand der Konkordanzen aus dem Korpus kann beispielsweise eine Richtig-Falsch-Unterscheidungsübung erstellt werden. Zusammen mit einem Lernerkorpus, bestehend aus Texten, die von den Schülern selbst verfasst wurden, könnten die Richtig-Falsch-Übungen erweitert und so das Verständnis der Schüler für muttersprachliches Spanisch geschärft werden. Auch die Korrektur eigener Texte kann auf diese Weise geübt werden. Die Schüler entwickeln ein besseres Verständnis für richtige Ausdrucksweisen und die Verbesserung ihrer eigenen Texte fällt ihnen folglich leichter. Allerdings müssen die Texte der Muttersprachler, die mit den Texten der Schüler verglichen werden, auch wirklich vergleichbar sein. Es sollten also beispielsweise Essays von muttersprachlichen Studenten mit den Aufsätzen der deutschen Schüler verglichen werden und keine Fachaufsätze, Romane oder Zeitungstexte. Ein Beispiel für die Erstellung einer Übung zur Fehlerkorrektur anhand eines Lernerkorpus ist in Kapitel 4.2.6. zu finden.

Mithilfe dieser Übungen, die mit realen Sprachdaten erstellt wurden, kann das sprachliche Bewusstsein der Schüler beträchtlich gefördert werden. Diese Förderung des sprachlichen Bewusstseins definiert James als eine Aktivität, die wiederum die Fähigkeit fördert, die Diskrepanz zwischen dem aktuellen Stand des eigenen Wissens und des

Zielniveaus zu erkennen und zu identifizieren.⁸²⁹ Auch Braun/Chambers geben an, dass mittels Korpora sprachliche Variationen in den Fremdsprachenunterricht gebracht werden können, wodurch vor allem bei fortgeschrittenen Lernern “das Verständnis für Sprache als ein komplexes und sich stetig wandelndes System” gefördert wird.⁸³⁰

In diesem Kontext ist die *Chemnitz Internet Grammar* (CING) von Hahn (2000) zu nennen. Es handelt sich dabei um ein computerbasiertes Programm, in dem Grammatikregeln, Übungen und reale Materialien in Form eines Korpus vereint sind und bei welchem die Rolle des Lehrers hinsichtlich Korrektur, Feedback und Unterrichtsführung zu einem gewissen Teil vom Computer übernommen wird. Diese lernerzentrierte, pädagogische Grammatik ist für das individuelle, selbstständige Erlernen des englischen Zeitensystems durch deutsche Schüler entworfen worden. CING bietet einen Dreifachansatz, indem es über eine regelbasierte Grammatik (mit grammatischen Erklärungen, Beispielen und Übungen), ein Korpus (das von den Lernern für eigene Recherchen verwendet werden kann) und eine interaktive Übungskomponente verfügt. Mittels dieses Programms haben die Lerner Zugang zu realen Daten, wodurch ein induktiver Lernprozess gefördert wird. Mithilfe der Internetgrammatik können aber auch deduktive und induktive Lernprozesse vereint werden. Zum einen kann der Schüler bei Fragen in der Referenzgrammatik nachsehen, was traditionell unter deduktivem Lernen verstanden wird. Zum anderen kann er auch mit dem Korpus (welches englische und deutsche Daten bereithält) arbeiten, beispielsweise bei der Betrachtung von Konkordanzen zum jeweiligen grammatikalischen Problem, und sich das Wissen somit induktiv aneignet. Auf diese Weise können unterschiedliche Lerner auch unterschiedliche Herangehensweisen wählen, was zu einer verstärkten Individualisierung der Lernwege führt und damit die Lernerautonomie fördert.⁸³¹

Cobb hat im Jahr 1997 den Lerneffekt durch die Verwendung von Konkordanzprogrammen untersucht. Dafür wurde ein experimenteller *Lexical Tutor* entwickelt, welcher den Lernern neue Lexeme auf unterschiedliche Weise vermittelt,

⁸²⁹ Vgl. James, Carl (1996): „A cross-linguistic approach to language awareness“. In: *Language Awareness* 5, S. 142.

⁸³⁰ Braun/ Chambers (2009): op. cit., S. 335.

⁸³¹ Vgl. Hahn, Angela (2000): „Grammar at its best: The development of a rule- and corpus-based grammar of English tenses“. In Burnard, L./ McEnery, T. (Hrsg.) (2000): *Rethinking language pedagogy from a corpus perspective: papers from the Third International Conference on Teaching and Language Corpora*. Frankfurt am Main: Peter Lang, S. 193-205.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

unter anderem in Konkordanzen. Dabei konnte ein beständiger Erfolg für die Einführung der Lexeme anhand von Konkordanzen erzielt werden.⁸³² Gaskell/Cobb nutzten im Jahr 2004 Konkordanzen, um das Wissen von Lernern in den ersten Lernjahren im Hinblick auf die Grammatik zu fördern. Sie konnten zeigen, dass die Lerner diese gerne verwenden wollen und sie dazu fähig sind, auf der Basis derselben Korrekturen durchzuführen, obgleich die Autoren zugeben, dass die Anpassung der Konkordanzen für die grammatikalische Entwicklung komplexer ist als für die lexikalische.⁸³³

Nach Braun/Chambers befassen sich die meisten Arbeiten zur didaktischen Nutzung von Korpora mit dem Einsatz konkordanzbasierter Materialien im Unterricht fortgeschrittener Lerner. Ihrer Meinung nach eignen sich Konkordanzen jedoch bereits im Sekundarschulbereich zur Veranschaulichung von Sprachverwendungsmustern. Teilweise veranschaulichen sie grammatische Muster besser als traditionelle Lehrwerke.⁸³⁴

3.3.1.2. *Teaching to exploit*

Das eben besprochene *Exploiting to teach* impliziert gewissermaßen automatisch *Teaching to exploit* im Schulunterricht, aber auch an der Universität. Sofern ein Schüler oder Student für bestimmte Untersuchungen Zugang zu einem Korpus hat, muss er zwangsläufig darin unterrichtet werden, wie man Korpora benutzt. Natürlich kann diese Konfrontation mit Korpora auch umgangen werden, indem den Schülern aus einem Korpus extrahierte Daten in Papierform vorgelegt werden. Diese Vorgehensweise wird zwar von einigen Wissenschaftlern mit der Begründung kritisiert, dass man es sich damit zu leicht mache, anstatt den Schülern das eigene Forschen im Korpus als Phase des *Teaching to exploit* nahezubringen, um damit auch bessere Langzeitergebnisse zu erzielen. Leech sieht diese Art der Vermittlung aber als vorteilhaft, da so, seiner Erfahrung nach, die maximale Anzahl an Schülern an dieser Art der Lernerfahrung teilnehmen können und wollen. Weder technikfeindliche noch interessierte Schüler, welche sich das notwendige Wissen zur Arbeit mit Korpora aneignen wollen, werden

⁸³² Vgl. Cobb, Tom (1997): „Is there any measurable learning from hands on concordancing?“ In: *System*, 25(3), S. 301–315.

⁸³³ Vgl. Gaskell, Delian/ Cobb, Tom (2004): „Can learners use concordance feedback for writing errors?“. In: *System*, 32, S. 301–319.

⁸³⁴ Vgl. Braun/ Chambers (2009): op. cit., S. 334.

dadurch entmutigt und alle können danach in ihrem eigenen Tempo und auf ihre eigene Art und Weise fortfahren.⁸³⁵

Teaching to exploit ist also der Unterricht über die Nutzungsmöglichkeiten von Korpora, es wird den Schülern praktisches Wissen an die Hand gegeben, damit sie die Korpora für ihre eigenen Zwecke nutzen können.⁸³⁶ Bei der Frage danach, wie Korpora im Unterricht verwendet werden sollten, scheint der Ansatz der *Three Is* (*Drei-I-Ansatz*: Illustration, Interaktion und Induktion) von Carter/McCarthy⁸³⁷ derzeit der am besten geeignete zu sein. *Illustration* bedeutet, dass reale Daten betrachtet werden, *Interaktion* heißt, Meinungen und Beobachtungen werden diskutiert und geteilt, und *Induktion* bezieht sich darauf, dass eigene Regeln für bestimmte Besonderheiten erstellt werden, welche dann weiterentwickelt und verfeinert werden je mehr Daten dazukommen. Dieser explorative Ansatz geht mit Johns Konzept des *Data-Driven Learning* (DDL) einher.⁸³⁸

Der DDL-Ansatz bezeichnet die Nutzung des Korpus als Informationsquelle für Schüler.⁸³⁹ Dieser Ansatz wurde im Jahr 1986 von Johns für den Unterricht der englischen Grammatik mittels der Analyse von Konkordanzen entwickelt.⁸⁴⁰ Im Zentrum des Sprachenlernens steht hier die eigene Entdeckung der Grammatik durch die Lerner, welche anhand von Belegen aus dem realen Sprachgebrauch stattfindet. Johns Meinung nach ist die Forschung zu bedeutend, um sie den Forschern zu überlassen. Daher argumentiert er, dass es das Ziel ist, die Lerner einer Sprache dazu zu ermutigen, ‚Forscher‘ zu werden, deren Lernprozess durch den Zugang zu linguistischen Daten angetrieben werden sollte. Im Sinne des DDL-Ansatzes führt die induktive Arbeit mit Korpora von der Betrachtung der Konkordanzen und Spekulationen seitens der Lerner über das Erkennen von Mustern in der Zielsprache schließlich zur Bildung von Verallgemeinerungen, welche diese Muster erklären. Johns identifiziert dabei drei Phasen: Identifizierung, Klassifizierung und Verallgemeinerung.⁸⁴¹ Laut McEnery/Xiao

⁸³⁵ Vgl. Leech (1997): op. cit., S. 10.

⁸³⁶ Vgl. McEnery/ Xiao (2010): op. cit., S. 370.

⁸³⁷ Vgl. Carter, Ronald/ McCarthy, Michael (1995): „Grammar and the spoken language“. In: *Applied Linguistics*, 16 (2), S. 141–158.

⁸³⁸ Vgl. McEnery/ Xiao (2010): op. cit., S. 370.

⁸³⁹ Vgl. Ciesielska-Ciupek (2003): op. cit., S. 522-523.

⁸⁴⁰ Vgl. Da Silva, Iandra Maria: *La utilización de la lingüística de corpus en la enseñanza de español: usos de pero y sino*. Online Publikation: http://cvc.cervantes.es/ensenanza/biblioteca_ele/publicaciones_centros/PDF/rio_2006/26_dasilva.pdf (06.06.2017), S. 222.

⁸⁴¹ Vgl. Johns (1991): op. cit., S. 2-4.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

stimmen Johns' induktiver Ansatz und Carter/McCarthy's Ansatz der *Three Is* grob überein.⁸⁴²

Für Johns ist der Computer kein Ersatz der Lehrkraft, sondern eine Art Informant. Der Unterschied zur bisherigen Nutzung von Computern zur Korrektur von sprachlichen Aussagen ist, dass er hier nicht als allwissendes Werkzeug zur Verbesserung von Schüleräußerungen eingesetzt wird, sondern als Quelle von sprachlichem Material. Es ist schließlich keinem Computer möglich, alle Regeln einer Sprache zu verinnerlichen, solange diese nicht vollständig erforscht ist, also komplett durch Regeln erklärt werden kann. Die Lerner sollen dadurch Zugang zur linguistischen ‚Performanz‘ erhalten und Johns traut ihnen zu, dass sie anhand der Daten aus dem KWIC-Format (siehe Kapitel 1.4.3.) nützliche Informationen als Antwort auf zuvor gestellte Fragen ziehen können. Während die Vorgehensweise des eigenständigen Entdeckens von Regeln bereits aus den traditionellen *rule-hiding-activities* bekannt ist, bei denen die Lehrkraft bereits weiß, auf welche Regeln die Übung hinauslaufen wird, ist hier nicht klar, welche Verallgemeinerungen die Schüler aus dem Korpus ziehen werden. Es können laut Johns auch Regeln entdeckt werden, die nicht nur dem Lehrer, sondern auch den gängigen Nachschlagewerken bislang unbekannt waren. Seiner Meinung nach gibt eben diese Herausforderung dem DDL-Ansatz seinen Reiz.⁸⁴³ Laut Leech vermittelt dieser autonome Lernprozess dem Schüler die realistische Erwartung, als ‚Forscher‘ Neuland zu betreten, indem er einen einzigartigen und individuellen Beitrag leistet, anstatt die Forschungen anderer erneut durchzuarbeiten und zu bewerten.⁸⁴⁴

Beim DDL-Ansatz handelt es sich um einen lernerzentrierten Lernprozess, bei dem der Lehrer eine Art Regisseur oder Koordinator ist. Für manche Lehrer ist es schwierig, sich daran zu gewöhnen, dass sie nicht mehr im Zentrum des Lernprozesses stehen und dass nicht klar ist, worauf die aktuelle Suche im Korpus hinausläuft. Man folgt in diesem Fall schließlich keinem Lehrplan und es gibt keine ‚Lösungsseite‘ wie diejenige zum Schluss des Lehrbuchs.⁸⁴⁵ Allerdings ist es wichtig zu wissen, merken McEnery/Xiao an, dass der Schlüssel zu erfolgreichem DDL – trotz der Bedeutung der Schülerzentrierung – der richtige Grad der Lehrerführung oder der pädagogischen Vermittlung ist. Dieser

⁸⁴² Vgl. McEnery/ Xiao (2010): op. cit., S. 371.

⁸⁴³ Vgl. Johns (1991): op. cit., S. 1-3.

⁸⁴⁴ Vgl. Leech (1997): op. cit., S. 10.

⁸⁴⁵ Vgl. Johns (1991): op. cit., S. 3.

hängt vom Alter der Lerner, deren Erfahrung sowie sprachlichem Niveau ab, denn ein Korpus ist nicht leicht zu bedienen. Es ist genauso einfach aus den Korpusbelegen unsinnige wie sinnvolle Schlussfolgerungen zu ziehen. Daher muss die Lehrkraft das notwendige Training in Bezug auf die Korpusanalyse durchlaufen.⁸⁴⁶ Der Lehrer ist bei dieser Art des Lernprozesses also ein Vermittler, dessen Ziel es ist, die Schüler mit linguistischen Korpora vertraut zu machen mittels einer lernerzentrierten Vorgehensweise, welche die Schüler dazu bringen soll, Fragen zu formulieren und diese mithilfe des Korpus zu beantworten. Dabei betont McCullough die Autonomie der Lerner in einer solchen Lernphase sowie die Neugier, die bei der Korpusarbeit seitens der Schüler vorhanden ist, da nie klar ist, was in einem Korpus gefunden wird.⁸⁴⁷ Zunächst betrachten Lehrer und Schüler also gemeinsam reale Daten, diskutieren darüber, tauschen ihre Meinung dazu aus und versuchen, zusammen Erklärungen für die Belege zu finden. Schüler und Lehrer ‚forschen‘ sozusagen im Korpus. Später, wenn die Vorgehensweise bekannt ist, können die Schüler durchaus selbstständig Suchen im Korpus durchführen.⁸⁴⁸

Die Schüler können somit nicht nur von der Lehrkraft geleitet, zuvor wohl definierte und ihnen genau vorgeschriebene Suchen in den Korpora durchführen, sondern die Korpora auch für ihre eigenen Zwecke nutzen, sozusagen als Ressource für autonomes Lernen. Sie sollten vor der eigenständigen Verwendung über Korpora informiert werden und die für deren Erstellung wichtigen Kriterien sowie die Schlussfolgerungen, die aus ihnen gezogen werden können, verstehen.⁸⁴⁹ Der Lehrer muss die Schüler vorher genauestens auf die Arbeit mit den Korpora vorbereiten, sie in die Funktionsweise, Möglichkeiten und Grenzen derselben einführen, ihnen Hilfestellungen geben und ihnen anhand einiger gemeinsamer Übungen ein Gefühl für die Korpora vermitteln. Er muss es ihnen ermöglichen, Erfahrung mit den Korpora zu sammeln, damit sie daraufhin bereit sind, selbst mit diesen zu arbeiten. Sofern diese Kriterien erfüllt sind, können die Schüler Antworten auf Fragen zu Lexik oder Grammatik selbstständig beantworten, indem sie das Korpus zurate ziehen, dort nachsehen, welche Formen verwendet werden, wie häufig diese vorkommen und dann entscheiden, ob diese Ergebnisse aussagekräftig und ob dies die richtigen Antworten auf ihre Fragen sind.

⁸⁴⁶ Vgl. McEnery/ Xiao (2010): op. cit., S. 370-371.

⁸⁴⁷ Vgl. McCullough (2001): op. cit., S. 132.

⁸⁴⁸ Vgl. Tolchinsky (2014): op. cit., S. 13-15.

⁸⁴⁹ Vgl. Aston (2000): op. cit., S. 7-8.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

Wichtig ist, dass die Schüler das Korpus als eine Möglichkeit der Recherche und des Lernens kennenlernen. Es ist wünschenswert, dass sie die Korpora nicht nur in der jeweiligen Unterrichtsstunde nutzen, sondern diese als Quelle realer Sprachausschnitte kennen und schätzen lernen, welche sie auch selbst für ihren eigenen Lernprozess nutzen können.

Zur Nutzung von Korpora durch Schüler im Allgemeinen ist anzumerken, dass es zwei Vorgehensweisen gibt, welche laut Hunston unterschiedliche Auswirkungen mit sich bringen. Zum einen kann der Lehrer zusammen mit den Schülern direkt am Korpus arbeiten, wodurch die Motivation der Schüler aufgrund der Arbeit mit realen Sprachdaten gesteigert wird, der Lehrer jedoch weniger Kontrolle hat. Ein Schüler hat beispielsweise eine dringende Frage, die geklärt werden muss, damit er seinen Aufsatz weiterschreiben kann. Hier kann es bisweilen vorkommen, dass zu Beginn der Korpusarbeit gestellte Fragen durch die Suche im Korpus nicht beantwortet werden. Es kann sein, dass die Lehrkraft sich bei dieser Vorgehensweise unsicher ist oder dass die mediale Ausstattung der Schule eine Nutzung der Korpora durch die Schüler im Unterricht nicht zulässt. Zum anderen kann der Lehrer das Thema und das zu untersuchende Material zuvor auswählen, was die Untersuchung für ihn zwar planbarer macht, jedoch zu einer Reduzierung der Motivation bei den Schülern führt. Der Vorteil bei dieser Vorgehensweise ist, dass die sprachlichen Belege ausgedruckt und den Schülern in Papierform vorgelegt werden können, sofern nicht genügend Computer zur Verfügung stehen.⁸⁵⁰ Für fortgeschrittene Lerner ist die erste Variante ganz klar vorzuziehen, da sie mit der Fremdsprache bereits vertraut sind und mit der Fülle an sprachlichem Material, obgleich sie vielem Unbekanntem begegnen, gut umgehen können. Korpora beinhalten allerdings auch Sätze, die Muttersprachler intuitiv als falsch einschätzen würden. Dies erkennen Schüler bei der selbstständigen Arbeit mit Korpora jedoch nicht, wodurch eine Bearbeitung mancher Daten, die Lenkung durch den Lehrer oder zumindest die Überprüfung anhand einer Grammatik manchmal sinnvoll erscheint.

Der Vorteil der zweiten Variante ist, dass sie bereits bei Anfängern eingesetzt werden kann. Die Analyse von zuvor von der Lehrkraft sorgfältig ausgewählten Konkordanzen zur Veranschaulichung von Vokabular oder zur Ermittlung neuer Grammatikregeln im Klassenverband motiviert bereits Lerner im Anfangsunterricht,

⁸⁵⁰ Vgl. Hunston (2002): op. cit., S. 171.

wenn sie selbst Neues herausfinden können. Vor allem ganz zu Beginn des Fremdsprachenunterrichts könnten die Konkordanzen auch einem selbst erstellten Korpus mit den bereits behandelten Lehrbuchtexten entnommen werden, sodass das Sprachmaterial den Schülern vertraut ist (siehe Kapitel 4.7.). Cobb/Horst beispielsweise führten ein Experiment durch, bei dem Englischschüler mit geringen Vokabelkenntnissen Lexeme anhand von Konkordanzen aus einem Korpus mit Lehrbuchtexten lernen sollten, um später einen Test zu absolvieren, der auf dem Cambridge Lexikon mit 2387 Lexemen basierte. In der ersten Aufgabe wurde nach Definitionen für 18 zufällig ausgewählte Lexeme gefragt. Die zweite Aufgabe war ein Lückentext aus zwei unbekannten Passagen, in welche sie 15 Lexeme einsetzen sollten. Die Autoren fanden dabei heraus, dass die Schüler erfolgreicher waren, wenn sie mit den Konkordanzen lernten als mit anderen Methoden, vor allem beim zweiten Aufgabentyp.⁸⁵¹

Für fortgeschrittene Lerner ist das sogenannte *Discovery Learning* vorteilhaft. Als eine Art des DDL sollen die Schüler hier dazu angehalten werden, ihre eigenen Untersuchungen im Korpus durchzuführen und dabei vom *Serendipity*-Effekt zu profitieren. Diese Vorgehensweise, bei welcher die Schüler sich sozusagen vom ‚glücklichen Zufall‘ durch das Korpus führen lassen und diejenigen sprachlichen Eigenheiten weiterverfolgen, auf die sie gerade stoßen, setzt allerdings ein nicht allzu striktes Unterrichtsprogramm voraus.⁸⁵²

Silvia Bernardini schreibt, dass das *Discovery Learning* (DL), auch genannt *The learner as traveller* einfach und erfolgversprechend an die Arbeit mit Korpora angepasst werden kann. Die Lerner sollen dabei unter der Leitung des Lehrers große und vielfältige Textsammlungen erforschen. Im Gegensatz zu Johns' (1991) Ansatz der Korpusarbeit als *Learning as research* hat Bernardini das *Learning as discovery* gewählt, weil hier nicht angenommen wird, dass die Lerner über dieselben Interessen, Kompetenzen und Kapazitäten verfügen wie Lehrer und Linguisten. Stattdessen sollen sie ihren eigenen Interessen folgen, während die Lehrkraft ihnen Gelegenheiten dazu bietet, ihre Kapazitäten und Kompetenzen weiterzuentwickeln, sodass ihre Suchanfragen besser fokussiert sind, ihre Ergebnisinterpretationen präziser sowie ihr Verständnis der

⁸⁵¹ Vgl. Cobb, Tom/ Horst, Marlise (2001): „Reading academic English: carrying learners across the lexical threshold“. In: Flowerdew, J./ Peacock, M. (Hrsg.) (2001): *Research Perspectives on English for Academic Purposes*. Cambridge: Cambridge University Press. S. 315-329.

⁸⁵² Vgl. Hunston (2002): op. cit., S. 171.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

Korpusnutzung und ihr Sprachbewusstsein schärfer werden. Ihr ist dabei durchaus bewusst, dass dieses Vorgehen zunächst verwirrend scheint, weil die Lerner tief verwurzelte Normen des Unterrichtsverhaltens aufgeben sollen, doch bald werden ihrer Meinung nach sowohl die Lehrkraft als auch die Schüler die Vorteile erkennen. Der Lehrer muss nicht immer die Quelle absoluten und grenzenlosen Wissens sein und die Schüler beginnen, sich selbst als aktive Teilnehmer des Lernprozesses zu sehen.⁸⁵³

Bernardini ist überzeugt vom *Serendipity*-Effekt und gibt an, dass Lerner nicht nur von der Lehrkraft geführt, sondern durchaus alleine in Korpora suchen und forschen können. Sie vertritt die mediationsorientierte Sichtweise von Bedeutung, wonach das Lernen vorwiegend eine Entwicklung der Kommunikations- und Lernkapazität ist, welche eingeübt und im Folgenden verstärkt und restrukturiert werden muss, um zu kommunikativer Kompetenz zu gelangen. Lernen ist ihrer Meinung nach vielmehr prozedural als deklarativ zu sehen und beruht auf drei Grundpfeilern. Zunächst einmal werden Ansätze bevorzugt, welche auf Aufgaben basieren, die die direkte Beschäftigung des Lerners mit der Aufgabe fordern. Zweitens sollen die Aufgaben zu differenzierten sprachlichen Äußerungen führen. Schließlich soll ein mediationsorientierter Ansatz den Lernern helfen, autonome Strategien in Bezug auf den Umgang mit Übungen zu entwickeln. Sie ist der Meinung, dass diese *pedagogy of discovery* für das Aushandeln von Bedeutung essentiell ist und die Schüler mithilfe dieser Art des aktiven und direkten Umgangs mit Aufgaben ihr Wissen erweitern sollten. Sie berichtet, dass ihre Schüler keineswegs verwirrt waren durch die mangelnde Lenkung seitens der Lehrkraft und dass diese ausnahmslos freiwillig mehr Zeit in die Korpusarbeit investierten als die ihnen vorgegebene. Des Weiteren sollen einige Schüler ihr mitgeteilt haben, sie hätten das Korpus (in diesem Fall das *British National Corpus*) auch für andere Sprach- und Übersetzungskurse verwendet, und sie gaben an, dass die Arbeit mit dem Korpus äußerst aufschlussreich für sie war.⁸⁵⁴

Einem solchen Ansatz folgend kann die Lehrkraft die Schüler, sofern sich bei diesen ein gewisser Grad der Selbstständigkeit eingestellt hat, alleine mit dem Korpus arbeiten lassen. Die Schüler starten beispielsweise bei einem interessanten Beispiel,

⁸⁵³ Vgl. Bernardini (2004): op. cit., S. 22-23.

⁸⁵⁴ Vgl. Bernardini, Silvia (2000): „Systematising serendipity: Proposals for concordancing large corpora with language learners“. In: Burnard, L./ McEnery, T. (Hrsg.) (2000): *Rethinking language pedagogy from a corpus perspective: papers from the Third International Conference on Teaching and Language Corpora*. Frankfurt am Main: Peter Lang, S. 227-229.

welches sie in der Schule oder bei ihren Hausaufgaben gelesen haben. Im Folgenden forschen sie völlig frei von Angaben ihres Lehrers in einem Korpus, finden die Bedeutung unbekannter Lexeme heraus, informieren sich über häufige Kollokationen (siehe Kapitel 4.1.1.) und ermitteln etwaige positive oder negative Konnotationen. In der nächsten Unterrichtsstunde können die Ergebnisse verglichen und besonders auffällige Verbindungen oder Zusammenhänge besprochen werden. Ein guter Ausgangspunkt für eine solche Suche kann eine Redewendung oder eine Kollokation sein, deren Bedeutung oder typischen Kontext die Schüler mithilfe des Korpus ermitteln sollen. Bernardini schreibt hierzu, dass die Motivation im Prozess der Aufgabe erhalten bleibt, wenn Lerner Korpora verwenden, um von ihnen selbst gestellte Fragen zu beantworten.⁸⁵⁵

Es ist auch denkbar, dass die Lehrkraft bei Korrekturen von Schülerarbeiten im vierten oder fünften Lernjahr die inkorrekten Passagen lediglich kennzeichnet und die Schüler selbst in einem Korpus nach der richtigen Formulierung suchen lässt. Diese Vorgehensweise ist beispielsweise bei Fehlern mit den Verben *ser/ estar/ hay* denkbar. Auch Braun/Chambers sehen Korpora als nützliches Hilfsmittel zur Überarbeitung von Texten, welche die Lehrkraft zuvor korrigiert hat.⁸⁵⁶

Wenn die Schüler nicht vollkommen frei in Korpora suchen sollen, ist es auch möglich, wichtiges Vokabular der Lektion oder des aktuellen Themengebiets von den Schülern im Korpus suchen und die häufigsten Kombinationsmöglichkeiten den anderen Schülern präsentieren zu lassen. Diese Übung scheint vor allem zur Vorbereitung im Hinblick auf die Textproduktion im Abitur, aber unter Umständen auch auf schriftliche und mündliche Schulaufgaben geeignet. Die Schüler könnten außerdem im Anschluss daran eine Mindmap zu diesem Thema oder Begriff und den damit häufig erscheinenden Lexemen erstellen. Generell kann der Lehrer auch Hypothesen zu bestimmten Themen aufstellen und die Schüler nach Beispielen zur Verifizierung suchen lassen.

Im Jahr 2004 analysierten Yoon/Hirvela Korpora als Hilfsmittel für Textproduktionen in zwei akademischen Schreibkursen für den Zweitspracherwerb. Sie interessierten sich dabei insbesondere für die Art der Korpusnutzung durch die Lerner sowie für die Wahrnehmung der Stärken und Schwächen dieses Hilfsmittels und fanden anhand qualitativer und quantitativer Daten heraus, dass die Lerner das Korpus für die

⁸⁵⁵ Vgl. Bernardini (2004): op. cit., S. 20.

⁸⁵⁶ Vgl. Braun/ Chambers (2009): op. cit., S. 335.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

Entwicklung der Schreibfähigkeit in der Zweitsprache nützlich fanden und sie mehr Vertrauen in ihre Fähigkeit zur Textproduktion hatten.⁸⁵⁷

Es ist gut vorstellbar, dass Lerner mithilfe von Korpora in Kombination mit traditionellen Nachschlagewerken wie Wörterbüchern oder Online-Übersetzungsprogrammen ihre schriftliche Ausdrucksfähigkeit ohne die Hilfe des Lehrers verbessern können. Auch McCullough ist der Meinung, dass Korpora und Konkordanzprogramme für das Schreiben von Texten und das Studium einer Sprache so nützlich sind wie automatische Rechtschreib- und Grammatikkorrekturprogramme, Wörterbücher und Thesauren.⁸⁵⁸ Eine Studie von Chambers/O'Sullivan zeigte, dass englische Lerner der französischen Sprache ein Korpus als nützliches Instrument zur Überarbeitung bereits von der Lehrkraft korrigierter Texte empfanden. So konnten Interferenzfehler, welche mithilfe von Lehrbüchern, Grammatiken und Wörterbüchern nur sehr schwer oder gar nicht selbst verbessert werden konnten, mit Korpora durchaus korrigiert werden. Die Ergebnisse der Studie suggerieren daher, dass die Nutzung eines geeigneten Korpus für Schüler eine probate Möglichkeit bietet, aktiv ihre Fähigkeiten zur Textproduktion weiterzuentwickeln, sofern die Lehrkraft sie darauf vorbereitet und ihnen helfend zur Seite steht.⁸⁵⁹

Laut Guy Aston bleibt die Frage zu klären, ob bei der Beschäftigung der Schüler mit den Korpora eine Mediation durch die Lehrkraft wirklich wünschenswert ist oder ob die Lerner nicht vorzugsweise selbst mit den Korpora arbeiten sollten. Der direkte Zugang zu Korpora und damit die eigenständige Arbeit mit und Analyse der Daten eröffnet dabei ganz neue Möglichkeiten für autonomes Sprachlernen. Allerdings birgt diese Selbstständigkeit auch Gefahren, über welche die Schüler vorher durch einführende Gespräche, Tipps und Warnungen informiert werden sollten. Aston geht sogar noch einen Schritt weiter. Er ist der Meinung, es könne den Lernern auch gezeigt werden, wie sie selbst ein eigenes Subkorpus aus Texten anderer Korpora erstellen und für die linguistische Analyse textspezifischer Eigenschaften oder für ihre eigenen Übersetzungen verwenden könnten. Das Wissen über die Kriterien der Korpuserstellung ermögliche es

⁸⁵⁷ Vgl. Yoon, Hyunsook/ Hirvela, Alan (2004): „ESL student attitudes toward corpus use in L2 writing“. In: *Journal of Second Language Writing*, 13, S. 257–283.

⁸⁵⁸ Vgl. McCullough (2001): op. cit., S. 140.

⁸⁵⁹ Vgl. O'Sullivan, Íde/ Chambers, Angela (2006): „Learners' writing skills in French: Corpus consultation and learner evaluation“. In: *Journal of Second Language Writing*, 15, S. 49–68.

den Lernern, eine kritische Haltung gegenüber den verwendeten Korpora einzunehmen und die Grenzen der Daten, auf welche sie ihre Interpretationen stützen, zu erkennen.⁸⁶⁰ Allerdings ist die Erstellung eines Korpus für Schüler keine leichte Aufgabe, weshalb es schwer zu realisieren sein wird, Schülern im Fremdsprachenunterricht ein derartiges Projekt aufzubürden. Vielmehr kann der Lehrer ein Korpus erstellen und zusammen mit den Schülern damit arbeiten beziehungsweise die Schüler darin forschen lassen. In einem speziellen Seminar in der Oberstufe oder an der Universität ist eine derartige Vorgehensweise jedoch durchaus vorstellbar.

Im spanischsprachigen Bereich finden sich zur Veranschaulichung von *Teaching to Exploit* nur wenige Studien. Zum einen ist die Studie von De Jacobi zu erwähnen, welche mit ihren brasilianischen Schülern den Gebrauch des spanischen *mismo* im Vergleich zum portugiesischen *mesmo* untersuchte. Vor dem Hintergrund, dass der Gebrauch von *mismo* auch für fortgeschrittene brasilianische Lerner des Spanischen ein Problem darstellt, ließ sie die beiden Formen in einem 10 Mio. Lexeme umfassenden spanischen Zeitungskorpus und einem portugiesischen Zeitungskorpus mit 30 Mio. Lexemen von den Schülern analysieren und ihre eigenen Schlüsse ziehen. Ihrer Meinung nach stellt der Gebrauch von Korpora im Sprachunterricht eine gute Alternative zu oder Ergänzung von Wörterbüchern und Grammatiken dar. Sie schätzte vor allem die induktive Vorgehensweise und die Kooperation, welche die Schüler bei der gemeinsamen Analyse an den Tag legen mussten, um zu einem Ergebnis zu kommen.⁸⁶¹

Zum anderen ist die bereits erwähnte Studie von Davies aus dem Aufsatz „Student use of large annotated corpora to analyze syntactic variation“ zu nennen. Dabei handelte es sich um einen Internetkurs, bei dem fortgeschrittene Lerner des Spanischen Phänomene untersuchten, von denen vermutet wurde, dass sie einer Variation zwischen Registern, zwischen Dialekten oder in Bezug auf eine sich gerade vollziehende Veränderung unterliegen. Die drei Hauptsätze von Korpora, welche von den Studenten genutzt wurden, waren das CdE:Orig (Gesamt) in der Version von 2008 (siehe Kapitel 2.1.1.3. und 2.2.1.3.), das CREA in Kombination mit dem CORDE (siehe Kapitel 2.1.1.1. und 2.2.1.1.) sowie das Internet mittels der Suchmaschinen von *Google* und *Google Groups*. Die Studenten fanden dabei heraus, dass jedes Korpus seine eigenen Stärken und

⁸⁶⁰ Vgl. Aston (2000): op. cit., S. 14-16.

⁸⁶¹ Vgl. De Jacobi, Claudia (2002): „Computadores, corpora y la enseñanza de español en cursos de letras“. In: *Anuario brasileño de estudios hispánicos*, 12, S. 29-44.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

Schwächen hat, welche es geschickt zu kombinieren gilt. Mit seiner Studie konnte Davies zeigen, dass auch Lerner, die relativ unerfahren in syntaktischen Untersuchungen sind, durch Unterstützung und Feedback Korpora nutzen, Hypothesen und Suchanfragen formulieren, Daten organisieren, Hypothesen bezüglich sprachlicher Variation bestätigen oder verwerfen und unter Umständen sogar Gründe für sprachliche Variation finden können.⁸⁶²

Die in diesem Unterkapitel beschriebenen direkten Verwendungsweisen sind bislang hauptsächlich auf die Hochschulbildung beschränkt und haben noch nicht Einzug in den allgemeinen Sprachunterricht, geschweige denn in den Fremdsprachenunterricht, gehalten.⁸⁶³ Aus diesem Grund werden im vierten Kapitel Vorschläge zur Verwendung von Korpora im Spanischunterricht gemacht, bei denen beide Nutzungsmöglichkeiten, also sowohl *Exploiting to teach* als auch *Teaching to exploit*, eingesetzt werden.

3.3.2. Indirekte Nutzung von Korpora

Wie bereits erwähnt, fallen in diesen Bereich laut Leech das Verlegen von Nachschlagewerken, die Entwicklung von Unterrichtsmaterialien sowie die Erstellung von Sprachtests.⁸⁶⁴ McEnery/Xiao erweitern diese Aufzählung zudem um die Entwicklung von Lehrplänen.⁸⁶⁵

3.3.2.1. Beitrag für die Erstellung von Wörterbüchern

Laut McEnery/Xiao haben Korpora das Verlegen von Nachschlagewerken revolutioniert. Zum einen können reale Verwendungsbeispiele innerhalb weniger Sekunden aus maschinenlesbaren Korpora entnommen werden. Zum anderen können Informationen über die Häufigkeit von Lexemen bei der Erstellung genutzt werden, wie beispielsweise im *Collins COBUILD English Dictionary* von HarperCollins oder im *Longman Dictionary of Contemporary English*, beide aus dem Jahr 1995.⁸⁶⁶ Endruschat/Schmidt-Radefeldt legen dar, dass traditionelle Wörterbücher in der Regel aufeinander aufbauen und die Auswahl der Lemmata sowie deren

⁸⁶² Vgl. Davies (2004): op. cit., S. 259-269.

⁸⁶³ Vgl. McEnery/ Xiao (2010): op. cit., S. 370.

⁸⁶⁴ Vgl. Leech (1997): op. cit., S. 5-6.

⁸⁶⁵ Vgl. McEnery/ Xiao (2010): op. cit., S. 367-368.

⁸⁶⁶ Vgl. McEnery/ Xiao (2010): op. cit., S. 365.

Bedeutungsbeschreibungen häufig aus anderen Wörterbüchern übernommen werden. Bei der korpusbasierten Erstellung von Wörterbüchern hängt die Auswahl hingegen von der Relevanz, der Aktualität und der Frequenz der Lexeme ab.⁸⁶⁷ Auf diese Weise verwenden sogar diejenigen das Produkt korpusbasierter Forschung, die angeben, noch nie von einem Korpus gehört zu haben, so Hunston.⁸⁶⁸ Außerdem ermöglichen die Annotationen der Korpora die Integration textueller und soziolinguistischer Metadaten zum Gebrauch des jeweiligen Ausdrucks in das Wörterbuch, wie beispielsweise Angaben zu Register und Genre des Texts oder zum Alter des Sprechers. Das Wortarten-Tagging erleichtert zudem die Einordnung polysemer oder homographischer Lexeme. Des Weiteren lässt ein Monitorkorpus kleinste Bedeutungs- oder Gebrauchsveränderungen erkennen und hilft somit, die Wörterbücher auf dem neuesten Stand zu halten. Dabei darf nicht vergessen werden, dass Korpora die Intuition einzelner Lexikographen ergänzen oder widerlegen und die Einträge so akkurater und verlässlicher machen.⁸⁶⁹ Durch die Verfügbarkeit von Korpora treten laut Braun/Chambers lexikalische und grammatische Sprachverwendungsmuster zutage, welche allein durch die Betrachtung einzelner Texte oder die Befragung von Muttersprachlern nicht erkannt werden können.⁸⁷⁰ Zudem bietet die korpusbasierte Vorgehensweise insbesondere bei Fachsprachen den Vorteil, dass konzeptuelle Definitionen tatsächlich von Experten stammen.⁸⁷¹

Susan Hunston sieht fünf Schwerpunkte in der Veränderung, die Korpora bei der Erstellung von Nachschlagewerken wie Wörterbüchern und Grammatiken mit sich gebracht haben: Frequenz, Kollokationen und Phraseologie, Variation, Authentizität sowie Lexik in der Grammatik.⁸⁷² Während bisherige Grammatiken, wie beispielsweise *A Comprehensive Grammar of the English Language* von Quirk et al. (1985), nicht korpusbasiert waren, stützt sich die *Longman Grammar of Spoken and Written English* von Biber et al. (1999) vollständig auf das *Longman Spoken and Written English Corpus*, das 40 Mio. Lexeme umfasst. Sie enthält nur reale Beispiele aus dem Korpus und befasst

⁸⁶⁷ Vgl. Endruschat, Annette/ Schmidt-Radefeldt, Jürgen (2008): *Einführung in die portugiesische Sprachwissenschaft*. Tübingen: Narr Francke Attempto, S. 153.

⁸⁶⁸ Vgl. Hunston (2002): op. cit., S. 96.

⁸⁶⁹ Vgl. McEnery/ Xiao (2010): op. cit., S. 365-366.

⁸⁷⁰ Vgl. Braun/ Chambers (2009): op. cit., S. 330.

⁸⁷¹ Vgl. Endruschat/ Schmidt-Radefeldt (2008): op. cit., S. 153.

⁸⁷² Vgl. Hunston (2002): op. cit., S. 96.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

sich mit Unterschieden im Register sowie in gesprochener und geschriebener Sprache.⁸⁷³ Des Weiteren beinhaltet sie detaillierte Statistiken zu Häufigkeit, Verteilung und Gebrauch einer großen Anzahl an grammatischen Strukturen.⁸⁷⁴ Die *Collins COBUILD Series* geht noch einen Schritt weiter und konzentriert sich auf die Lexik in grammatischen Beschreibungen. In der sogenannten *pattern grammar* wird der Unterschied zwischen Lexik und Grammatik abgelehnt, was für den Sprachunterricht von Vorteil ist, da das Vokabular nicht isoliert betrachtet wird, sondern als Teil einer Phrase.⁸⁷⁵ Auch in der Romanistik hat die Korpuslinguistik zu einer tiefgreifenden Veränderung der Grammatikschreibung geführt. In der Hispanistik ist laut Klöden seit den 1990er Jahren eine rege Grammatikproduktion zu verzeichnen, welche die Besonderheiten der gesprochenen Sprache aufnimmt, wie beispielsweise die Grammatiken von Butt/Benjamin (1988)⁸⁷⁶ oder Bosque/Demonte (1999)⁸⁷⁷ zeigen.⁸⁷⁸

Korpora können nicht nur bei der Erstellung neuer Nachschlagewerke behilflich sein, sondern auch der Überprüfung bereits bestehender Nachschlagewerke und Sekundärliteratur dienen. So lässt Burr seit dem Jahr 1990 Studenten der Romanistik in ihren Seminaren anhand des von ihr selbst erstellten *Korpus romanischer Zeitungssprachen* computergestützte Korpusuntersuchungen zu bestimmten Grammatikbereichen durchführen, um ihnen Methoden sowie Einblicke in die tatsächliche Funktionsweise von Sprache zu vermitteln. Eine ihrer Studentinnen stellte dabei die Darstellungen der Tempora *perfecto compuesto* und *perfecto simple* (auch: *perfecto indefinido*) in der Sekundärliteratur deren Vorkommen im Korpus gegenüber. Sie fand heraus, dass der Tempusgebrauch im Hinblick auf die Verwendung von Zeitangaben wie *ayer*, *anoche*, *esa misma noche* oder auch *ahora* und *este año* tatsächlich demjenigen in der Sekundärliteratur entspricht. Zudem ermittelte sie unter anderem, dass im fortlaufenden Text häufiger das *perfecto simple*, in der wörtlichen Rede häufiger das *perfecto compuesto* steht.⁸⁷⁹

⁸⁷³ Vgl. McEnery/ Xiao (2010): op. cit., S. 365-366.

⁸⁷⁴ Vgl. Davies/ Face (2006): op. cit., S. 133.

⁸⁷⁵ Vgl. McEnery/ Xiao (2010): op. cit., S. 365-366.

⁸⁷⁶ Vgl. Butt, John/ Benjamin, Carmen (1988): *A New Reference Grammar of Modern Spanish*. London: Arnold.

⁸⁷⁷ Vgl. Bosque, Ignacio/ Demonte, Violeta (Hrsg.) (1999): *Gramática descriptiva de la lengua española*. 3 vols. Madrid: Espasa.

⁸⁷⁸ Vgl. Klöden (2002): op. cit., S. 9-10.

⁸⁷⁹ Vgl. Burr (2004): op. cit., S. 152-155.

Laut Siepmann werden in Schulgrammatiken oft seltene Ausnahmen vermittelt, zusammen mit der Annahme, sie seien beliebig einsetzbar, wenn man nur ihre Form beherrsche.⁸⁸⁰ Für Kaltenböck/Mehlmauer-Larcher ist die Häufigkeit jedoch kein ausschlaggebendes Kriterium für die Vermittlung von Ausdrücken im Unterricht, da häufige Lexeme und Grammatikeinheiten ihrer Meinung nach von Lernern ohnehin unterbewusst gelernt werden.⁸⁸¹ McEnery/Xiao erwidern hierauf jedoch, dass die Unterscheidung zwischen dem bestimmten und dem unbestimmten Artikel im Englischen, trotz deren häufigen Vorkommens, von chinesischen Schülern nicht nebenbei gelernt wird, da die Unterscheidung zwischen zählbaren und nicht-zählbaren Substantiven in deren Sprache nicht existiert.⁸⁸²

In Bezug auf den Englisch- und Französischunterricht an deutschen Schulen gibt es bereits einige Studien in diesem Bereich. Schlüter beispielsweise untersuchte das englische Perfekt mittels einer Korpusanalyse. Anhand seiner Studie zum Gebrauch des *present perfect* in der britischen und amerikanischen Gegenwartssprache konnte er alle Bedeutungsvarianten mit Häufigkeitsangaben versehen und damit zentrale von marginalen Funktionen abgrenzen. Er ermittelte durch seine Untersuchung auch einige neue linguistische Aspekte und konnte Schlussfolgerungen für die Vermittlung des *present perfect* im Englischunterricht ziehen.⁸⁸³ Mindt führte eine korpuslinguistische Studie zum Gebrauch der englischen Modalverben durch⁸⁸⁴, welche er im Jahr 2000 zu einer vollständigen Beschreibung des englischen Verbalsystems ausbaute.⁸⁸⁵ Für den Französischunterricht entwickelte Schmidt auf der Grundlage von Frequenzdaten ein Wörterbuch des intransparenten französischen Wortschatzes. Mithilfe dieses Vokabulars kann der deutsche Lerner des Französischen sich einen rezeptiven Wortschatz von 20.000 Lexemen aufbauen.⁸⁸⁶

⁸⁸⁰ Vgl. Siepmann (2009): op. cit., S. 325.

⁸⁸¹ Vgl. Kaltenböck, Gunther/ Mehlmauer-Larcher, Barbara (2005): „Computer corpora and the language classroom: On the potential and limitations of computer corpora in language teaching“. In: *ReCALL*, 17, S. 78.

⁸⁸² Vgl. McEnery/ Xiao (2010): op. cit., S. 374.

⁸⁸³ Vgl. Schlüter, Norbert (2002): *Present Perfect. Eine korpuslinguistische Analyse des englischen Perfekts mit Vermittlungsvorschlägen für den Sprachunterricht*. Tübingen: Narr.

⁸⁸⁴ Vgl. Mindt, Dieter (1995): *An empirical grammar of the English verb: modal verbs*. Berlin: Cornelsen.

⁸⁸⁵ Vgl. Mindt, Dieter (2000): *An empirical grammar of the English verb system*. Berlin: Cornelsen.

⁸⁸⁶ Vgl. Schmidt, Ulrike (2005): *Wortschatz zum Verstehen: Französisch. Nützliches Wörterbuch für fortgeschrittene Lerner*. Aachen: Shaker.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

In Bezug auf die Frequenz ist es Hunston zufolge, abgesehen von einigen allgemeinen Aussagen, fast unmöglich sich der relativen Häufigkeit von Lexemen, Phrasen und Strukturen bewusst zu sein. Beispielsweise sei es nicht schwer zu erraten, dass das englische Verb *take* häufiger gebraucht würde als das Verb *disseminate*. Schwieriger sei es, herauszufinden, ob *fare* oder *fantasy* häufiger verwendet würde.⁸⁸⁷ Auch Halliday vertritt diese Meinung und betont, dass Sprecher einer Sprache zwar intuitiv über die Häufigkeit einzelner Lexeme Auskunft geben könnten, ihnen jedoch weniger bewusst sei, welche grammatischen Kategorien sie benutzten. Das Niveau der Bewusstheit und Kontrolle läge in diesem Bereich viel tiefer, weshalb sich Muttersprachler über den Gebrauch von Aktiv und Passiv nicht so bewusst seien wie über die Verwendung einzelner Lexeme.⁸⁸⁸ Muttersprachler können somit allgemeine Informationen über die relative Häufigkeit eines Lexems im Vergleich zu anderen treffen. Ihre Aussagen sind allerdings nur intuitiv und daher nicht zuverlässig genug, um auf deren Basis Auskunft über die genaue Frequenz zu geben. Im Gegensatz dazu stellen Korpora eine Möglichkeit dar, objektiv Informationen über die Häufigkeit von Lexemen zu erhalten. Dies ist insbesondere der Vorteil bei der Arbeit mit repräsentativen Korpora als Sammlungen realen Sprachgebrauchs.⁸⁸⁹

Della Summers schreibt in ihrem Aufsatz “Computer lexicography: the importance of representativeness in relation to frequency” Folgendes über die Arbeit und die Ergebnisse der Korpuslinguisten:

It is commonly opined by corpus workers that although things that are discovered from corpus analyses are ‘obvious’, they only become obvious once the corpus has revealed them to us, i.e. they are not reliably recovered from the lexicographer’s innate knowledge and understanding of the language.⁸⁹⁰

Dieses Bewusstwerden wird von James/Garrett als “explicitmaking of implicit knowledge” bezeichnet.⁸⁹¹ Dank der Verfügbarkeit repräsentativer Korpora wie des

⁸⁸⁷ Vgl. Hunston (2002): op. cit., S. 21.

⁸⁸⁸ Vgl. Halliday, M. A. K. (1993): „Quantitative studies and probabilities in grammar“. In: Hoey, M. (Hrsg.) (1993): *Data, Description, Discourse: Papers on the English Language in honour of John McH. Sinclair*. London: HarperCollins, S. 3.

⁸⁸⁹ Vgl. Davies/ Face (2006): op. cit., S. 132.

⁸⁹⁰ Summers, Della (1996): „Computer lexicography: the importance of representativeness in relation to frequency“. In Thomas, J./ Short, M. (Hrsg.) (1996): *Using Corpora for Language Research*. London: Longman, S. 263.

⁸⁹¹ James, Carl/ Garrett, Peter (1992): *Language awareness in the classroom. Applied linguistics and language study*. London: Longman, S. 19.

Cobuild Corpus (Sinclair 1987), des *Longman Corpus* (Biber et al. 1999) und des *British National Corpus* (Burnard 2000) genießen Lerner des Englischen schon seit mehreren Jahren ein großes Angebot an exzellenten Lehrmitteln, die auf realen Daten aus repräsentativen Korpora basieren.⁸⁹² Bereits im Jahr 1987 wurde, wie bereits erwähnt, das *Collins COBUILD English Learner Dictionary* (Sinclair 1987) als erstes vollkommen korpusbasiertes Lernerwörterbuch veröffentlicht. Daraufhin folgten die anderen Verleger englischer Wörterbücher diesem Beispiel und ab dem Jahr 1995 beruhten alle großen Lernerwörterbücher im englischsprachigen Raum mehr oder weniger auf Korpusdaten.⁸⁹³ Beispiele hierfür sind das *Longman Dictionary of Contemporary English* (2003) und das *Cambridge Advanced Learner's Dictionary* (2005).⁸⁹⁴

Wie bereits zu sehen war, gab es im Bereich des Spanischen jedoch bis vor ein paar Jahren lediglich eine geringe Anzahl an großen, repräsentativen Korpora und damit auch wenige, wenn überhaupt wirklich repräsentative Frequenzwörterbücher.⁸⁹⁵ Dennoch wurden über die Jahre ein paar Häufigkeitswörterbücher entwickelt, wie beispielsweise Buchanan (1927), Eaton (1940), Rodríguez Bou (1952), García Hoz (1953), Juilland/Chang-Rodríguez (1964), Alameda/Cuetos (1995) sowie Sebastián/Carreiras/Cuetos (2000). Allerdings haben all diese Frequenzwörterbücher zumindest einen Makel im Hinblick auf die Entwicklung von Lehrmaterialien, so Davies/Face. Zum einen seien die meisten von ihnen mehr als 40 Jahre alt und repräsentierten daher nicht das aktuelle Spanisch. Zum anderen beschäftigten sich alle diese Wörterbücher nur mit geschriebenem, nicht aber mit gesprochenem Spanisch. Des Weiteren umfassten sie nur eine geringe Anzahl von Lexemen, in jedem Fall unter drei Mio. Einträgen. Alameda/Cuetos (1995) ist zwar relativ aktuell, listet allerdings keine Lemmata auf, sondern lediglich Wortformen, und basiert nur auf wenigen Texten, die außerhalb Spaniens entstanden sind. Sebastián/Carreiras/Cuetos (2000) ist lediglich in elektronischer Form vorhanden und außerhalb Spaniens nur äußerst schwer zugänglich.⁸⁹⁶

Im Jahr 2001 veröffentlichte die *Real Academia Española* dann zwei repräsentative Korpora – das CREA und das CORDE (siehe Kapitel 2.1.1.1. und 2.2.1.1.). Allerdings

⁸⁹² Vgl. Davies/ Face (2006): op. cit., S. 133.

⁸⁹³ Vgl. McEnery/ Xiao (2010): op. cit., S. 366.

⁸⁹⁴ Vgl. Davies/ Face (2006): op. cit., S. 133.

⁸⁹⁵ Vgl. Davies/ Face (2006): op. cit., S. 133.

⁸⁹⁶ Vgl. Davies (2005): op. cit., S. 106-107.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

konnten auf der Basis dieser beiden Korpora keine Frequenzwörterbücher für den Sprachunterricht erstellt werden, da sie, zumindest zum damaligen Zeitpunkt, nicht nach Lemmata und Wortarten getaggt waren. Somit konnten keine Aussagen über die Häufigkeit der Lemmata getroffen werden, wodurch die Erstellung eines Frequenzwörterbuchs wiederum scheitern musste.⁸⁹⁷

Ein Jahr später veränderte sich die Situation entscheidend mit der Veröffentlichung des CdE:Orig. Dieses Korpus ist, wie in Kapitel 2.1.1.3. bereits erwähnt, lemmatisiert sowie nach Wortarten getaggt und erlaubt es den Entwicklern von Frequenzwörterbüchern und korpusbasierten Lehrmitteln, nach Lemmata zu suchen und ambige Wortformen anhand der Wortart zu disambiguieren. Außerdem wurden das Taggen und die Lemmatisierung in den Jahren 2003 und 2004 zunehmend verbessert, wodurch das Korpus zu einer akkuraten und zuverlässigen Quelle für Frequenzwörterbücher wurde. Zudem ist das Korpus ungefähr 40 mal so groß wie das Korpus, auf welchem das Wörterbuch von Chang-Rodríguez (1964), welches lange Zeit als eines der besten spanischen Frequenzwörterbücher galt, basierte (allerdings befand man die Wortauswahl schon damals als veraltet).⁸⁹⁸

Von 2004 bis 2005 wurden die Informationen aus dem CdE:Orig für die Entwicklung eines Frequenzwörterbuchs verwendet, welches im Jahr 2005 unter dem Titel *A Frequency Dictionary of Spanish: Core Vocabulary for Learners* von Mark Davies veröffentlicht wurde.⁸⁹⁹ In Bezug auf die Möglichkeiten der Erstellung eines solchen Frequenzwörterbuchs gesteht Davies seinen Vorreitern auf dem Gebiet der Wörterbucherstellung einige Benachteiligungen zu. Seiner Meinung nach hätten erst die modernen technischen Möglichkeiten die Erstellung seines Frequenzwörterbuchs ermöglicht. Bei Juilland/Chang-Rodríguez (1964) sei ein solches Unterfangen schlichtweg unmöglich gewesen. Außerdem habe das Taggen des Korpus zwar mehrere Wochen gedauert, diese Zeitspanne sei allerdings wesentlich kürzer gewesen als noch vor fünf oder zehn Jahren.⁹⁰⁰

Allerdings sollten bei der Auswahl der Einträge in einem Wörterbuch nicht nur die Frequenzanalysen eines Korpus beachtet werden, sondern es ist auch notwendig nach

⁸⁹⁷ Vgl. Davies/ Face (2006): op. cit., S. 134.

⁸⁹⁸ Vgl. Davies/ Face (2006): op. cit., S. 134.

⁸⁹⁹ Vgl. Davies/ Face (2006): op. cit., S. 135.

⁹⁰⁰ Vgl. Davies (2005): op. cit., S. 115.

dem Nutzen eines Lexems für die Alltagskommunikation der Schüler zu fragen. So erscheinen einige Lexeme aufgrund ihrer Zugehörigkeit zum mündlichen Bereich oder zur Umgangssprache nur selten in Korpora und würden daher eigentlich nicht in das Wörterbuch aufgenommen werden. Hier muss die Entscheidung letztendlich vom Lexikografen getroffen werden. Diese Ansicht spiegelt sich auch in der Aussage von Della Summers wider, die zum Abschluss ihres Aufsatzes über die Erstellung von Wörterbüchern unter Zuhilfenahme von Korpora schreibt:

“However, in dictionary-making editorial judgment is of paramount importance, because blindly following a corpus, no matter how carefully it may be constructed to represent the target language type accurately, can lead to oddities. We expect our motto: ‘Corpus-based, but not corpus-bound’ to hold good for many years to come.”⁹⁰¹

Korpora können durch die Ermittlung der Häufigkeit von Lexemen aber nicht nur bei der Auswahl der Lexeme helfen, die in ein Wörterbuch integriert werden sollen, sondern auch bei der Anordnung der darin aufgelisteten Bedeutungsvarianten. So werden die einzelnen Bedeutungen oder Phraseologismen (siehe Kapitel 4.1.2.), in denen das jeweilige Lexem vorkommt, im *Longman Dictionary of Contemporary English* (Third Edition, 1995) nach deren Häufigkeit im *Longman Lancaster Corpus* und im *British National Corpus* angeordnet. Diejenige Bedeutung wird zuerst genannt, die im Korpus am häufigsten vorkommt, auch wenn diese nur Teil eines Phraseologismus ist.⁹⁰²

Im Hinblick auf die korpusbasierte Erstellung eines Wörterbuchs unter semantischem Blickwinkel ist auch das *Diccionario REDES* zu nennen.⁹⁰³ Das REDES ist ein Wörterbuch, welches auf einem Korpus mit 250 Mio. Lexemen beruht, allerdings handelt es sich dabei um ein Zeitungskorpus, kein Referenzkorpus. Das Korpus besteht aus Texten von 68 Zeitungen der spanischen und lateinamerikanischen Presse der letzten 20 Jahre. Das REDES ist das erste sogenannte *diccionario combinatorio* des Spanischen und soll eine Art Brücke zwischen der Lexikographie und der Grammatik darstellen, indem es die Kluft zwischen der Bedeutung eines Lexems und der Form, in der die Sprecher es gewöhnlich nutzen, verringert. Es unterscheidet sich von anderen Wörterbüchern darin, dass es die Lexeme nicht definiert, sondern Kontexte zeigt, in denen sie erscheinen. Dabei handelt es sich nicht um eine Sammlung von Wortlisten,

⁹⁰¹ Summers (1996): op. cit., S. 266.

⁹⁰² Vgl. Summers (1996): op. cit., S. 263-264.

⁹⁰³ Vgl. Bosque, Ignacio (2004): *REDES. Diccionario combinatorio del español contemporáneo*. Madrid: Ediciones SM.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

sondern um eine Zusammenstellung von Verbindungen zwischen Lexemen in Form semantischer Beziehungen und Restriktionen, die das Lexem explizit beschreiben. Für das Lexem *planear* wird beispielsweise angegeben, dass es im Spanischen nicht nur in Zusammenhang mit *aviones* oder *gaviotas* gebraucht wird, sondern auch mit *sombras*, *dudas*, *sospechas*, *peligros* oder *amenazas*. Es soll sowohl für spanischsprachige als auch nicht-muttersprachliche Schüler, Studenten, Lehrkräfte und Sprachwissenschaftler, Übersetzer, Schriftsteller sowie Spezialisten auf dem Gebiet der automatischen Sprachverarbeitung von großem Nutzen sein.⁹⁰⁴

Zwei Jahre später veröffentlichte der Herausgeber des REDES, Ignacio Bosque, das *Diccionario combinatorio práctico del español contemporáneo*, welches er auch mit der Kurzform PRÁCTICO betitelt. Es wurde ausgehend von REDES und dessen Zeitungskorpus entwickelt und dient wie dieses dazu, dem Nutzer den tatsächlichen Sprachgebrauch zu vermitteln. Das PRÁCTICO enthält 400.000 gewöhnliche Wortverbindungen, ist ideal, um den Wortschatz zu erweitern und die Ausdrucksweise flüssiger und präziser zu gestalten.⁹⁰⁵ Das PRÁCTICO ist ein Gebrauchswörterbuch. Während das REDES 8000 Lexeme mit ungefähr 200000 Kombinationen enthält, sind es im PRÁCTICO 14000 Einträge mit fast 400000 Verbindungen. Es enthält jedoch weniger Beispiele als das REDES, welche zudem nicht aus Texten entnommen, sondern von den Redakteuren entwickelt worden sind.⁹⁰⁶ Bosque veranschaulicht das Verhältnis der beiden Wörterbücher wie folgt:

„Podría decirse, en resumen, que REDES se parece a un mapa, mientras que PRÁCTICO está más cerca de ser una guía. Ciertamente, los mapas pueden usarse como guías, pero también pueden no usarse en absoluto y concebirse como descripciones más o menos detalladas del terreno. Las guías, por el contrario, aportan menos información que los mapas en relación con algunas cuestiones, pero más en cambio en lo relativo a otras.”⁹⁰⁷

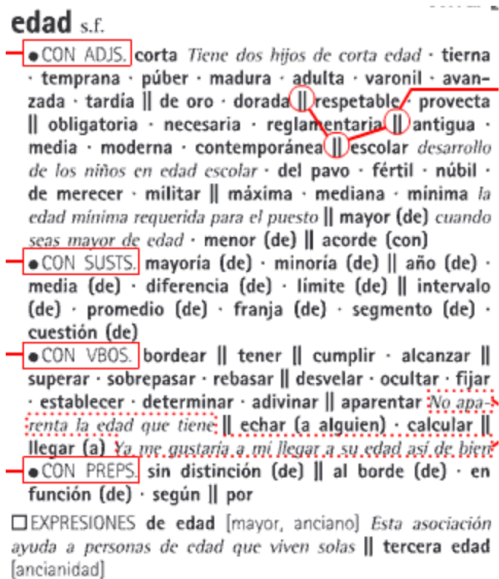
Zur Veranschaulichung wird der Eintrag des femininen Substantivs *edad* abgebildet:

⁹⁰⁴ Vgl. Bosque, Ignacio (2004): *REDES. Diccionario combinatorio del español contemporáneo. Las palabras en su contexto*. Online Publikation: http://www.rae.es/sites/default/files/PDF_diccionario_redes.pdf (01.10.2017).

⁹⁰⁵ Vgl. Grupo SM (ed.): *Las palabras en su contexto. Diccionario combinatorio práctico del español contemporáneo*. Online Publikation: <http://diccionariopractico.grupo-sm.com> (03.10.2017).

⁹⁰⁶ Vgl. Bosque, Ignacio (2006): *Diccionario combinatorio práctico del español contemporáneo. Las palabras en su contexto*. Madrid: Ediciones SM, S. XIV.

⁹⁰⁷ Bosque (2006): op. cit., S. XIV.

Abbildung 76: Ausschnitt aus dem *Diccionario combinatorio práctico del español contemporáneo*.⁹⁰⁸

Im PRÁCTICO wird jeweils angegeben, mit welchen anderen Wortarten der Ausdruck verwendet wird, sowie die üblichsten Beispiele dafür aufgelistet. Wortkombinationen werden durch einen Doppelstrich getrennt, Beispiele werden kursiv geschrieben.⁹⁰⁹ Die Korpusanalyse wurde somit dazu genutzt, besonders herausragende Kombinationen zu ermitteln, um das Wörterbuch so benutzerfreundlich wie möglich zu gestalten.

Die spanische Lexikographie muss allerdings laut Alonso-Ramos vor allem im Bereich der Lernerwörterbücher weiter vorangebracht werden. Sie führt das *Macmillan English Dictionary* an, welches in sogenannten ‚Get it right‘-Boxen Beispiele aus dem *International Corpus of Learner English* (ICLE) enthält. Hier wird beispielsweise für das Substantiv *need* angegeben, dass es mit der Präposition *for* und nicht *of* gebraucht wird, sofern darauf ein weiteres Substantiv folgt. Außerdem sind ein richtiges und ein falsches Beispiel darunter angegeben. Alonso-Ramos bedauert, dass bislang kein ähnliches Wörterbuch für die spanische Sprache erstellt wurde. Hierfür würde jedoch ein

⁹⁰⁸ Orihuela, José Luis (2006): *Diccionario combinatorio práctico del español contemporáneo*. Online Publikation: <https://www.ecuaderno.com/2006/12/07/diccionario-combinatorio-practico-del-espanol-contemporaneo/> (13.10.2017).

⁹⁰⁹ Vgl. Bosque (2006): op. cit., S. XVII.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

umfassendes Lernerkorpus (siehe Kapitel 2.2.2.) benötigt, welches nach Fehlern getagged sein müsste.⁹¹⁰

Für Sonntag können Korpora uns zudem bei unterschiedlichen Lesarten lexikalisierten Einheiten helfen, um anhand der jeweiligen Häufigkeit zu ermitteln, welchen Status diese haben. Er analysierte und taggte die lexikalisierten Formen auf *-illa* in einem selbst zusammengestellten Korpus mit Daten aus den Korpora CREA und CORDE und verglich diese mit deren Einträgen in Wörterbüchern. Neben der Ermittlung des Status der jeweiligen Einheiten geben die Informationen aus dem Korpus uns seiner Meinung nach auch Auskunft über den Wortschatz und seine aktuelle Entwicklung, beispielsweise, wenn Lesarten im Korpus, aber nicht im Wörterbuch vorkommen. Als Beispiele führt er *plantilla* an, welches mit der Lesart ‚Tastaturschablone‘/ ‚Dokumentvorlage‘ nur in sehr aktuellen Wörterbüchern vorkommt sowie *astilla*, welches mit der Lesart ‚Bestechungsgeld‘ in keinem ihm bekannten Wörterbuch verzeichnet ist. Im Falle des Ausbaus dieser Datensammlung sieht Sonntag die Möglichkeit, Einsichten in Diskrepanzen zwischen lexikographischer Information und der Sprachwirklichkeit, welche durch das Korpus zugänglich ist, zu gewinnen. Durch *tags* wie das Herkunftsland oder Inhaltsrelationen wie das Attribut ‚Metapher‘ können alle Beispiele, auf die das Kriterium zutrifft, aufgelistet werden, was die Untersuchung der Bedeutung im jeweiligen Land oder im jeweiligen Kontext ermöglicht (für weitere Informationen zum *tagging* siehe Kapitel 1.4.5.). Für die Zukunft erwartet sich Sonntag empirische Daten zur Überprüfung sprachwissenschaftlicher Thesen sowie Hilfe bei Begriffsbestimmungen beispielsweise hinsichtlich Polysemie, semantischer Felder, Lexikalisierung, Wortbildungsfunktionen und Bedeutungswandel. Zudem erhofft er sich Einsichten in die Fülle des spanischen Wortschatzes und in das Funktionieren dieses Wortschatzes im Textzusammenhang sowie Präzisierungen im Hinblick auf die diatopische Reichweite der Varianten.⁹¹¹

⁹¹⁰ Vgl. Alonso-Ramos, Margarita (2016): „Spanish learner corpus research: Achievements and challenges“. In: Alonso-Ramos, M. (Hrsg.) (2016): *Spanish Learner Corpus Research: Current trends and future perspectives*. Amsterdam [u.a.]: John Benjamins, S. 16-17.

⁹¹¹ Vgl. Sonntag, Eric (2007): „Diminutivische Sachbezeichnungen. Lexikalisierte Formen auf *-illa* zwischen Lexikographie und Korpuslinguistik“. In: Bierbach, M./ von Gemmingen, B./ Stork, Y. (Hrsg.) (2007): *Das gefesselte Wort. Beiträge zur Entwicklung von Wörterbüchern und Grammatiken des Spanischen*. Bonn: Romanistischer Verlag, S. 190-193.

In diesem Zusammenhang ist auch Buckingham's Studie zu nennen. Sie untersuchte *Construcciones con verbo soporte (CVS)* anhand eines 21 Mio. Lexeme umfassenden Korpus des geschriebenen akademischen Gegenwartsspanischen in Spanien und Lateinamerika. Sie analysiert in ihrer Studie CVS mit den Verben *dar*, *poner*, *tener* und *tomar* und berücksichtigt dabei die Struktur und das stilistische Niveau der Anwendung der CVS sowie das Niveau der strukturellen und semantischen Äquivalenz, welche die spanischen Konstruktionen mit ähnlichen Konstruktionen im Englischen und Deutschen haben. Sie hat dabei auch eine didaktische Zielsetzung: Ihrer Meinung nach können sich Lerner der spanischen Sprache diese CVS nicht adäquat aneignen, da sie in Wörterbüchern nicht systematisch behandelt werden. Daher macht sie, basierend auf den morphosyntaktischen Informationen aus dem Korpus, Vorschläge für ein Modell lexikografischer Einträge dieser CVS.⁹¹²

Eine systematische Behandlung der CVS in Schulwörterbüchern ist sicherlich die Voraussetzung dafür, dass Schüler sich diese adäquat aneignen können. Auch die von Sonntag angesprochene Häufigkeit der Lexeme ist für Schulwörterbücher nicht zu unterschätzen. Zwar ist sie allein nicht das ausschlaggebende Kriterium, doch sie sollte bei der Auswahl und Anordnung durchaus beachtet werden. Ebenso scheint eine frequenzbasierte Auflistung der Einzelbedeutungen innerhalb eines Eintrags, wie im Fall der *Longman Dictionaries*, sinnvoll, da die Schüler somit die häufigste Bedeutung gleich zu Beginn des Eintrags finden würden und beim Erlernen des Vokabulars sofort wüssten, welche Bedeutungen sie vermutlich am häufigsten brauchen werden.

Davies gibt an, dass man für das Englische in der Studie *National* (2000) herausfand, man könne mit den 1000 häufigsten Lexemen des Englischen bereits 80% aller Lexeme in einem Text verstehen. Mit den 1000 nächsthäufigsten Lexemen sind dem Lerner 5% mehr Lexeme der Texte bekannt und die Rangplätze 2000 bis 3000 erweitern den Wortschatz nur noch um 3-4%. Solche Daten sind durchaus von Bedeutung für den Fremdsprachenlehrer und -lerner, da mit der Auswahl des richtigen Lehrwerk vokabulars das Verständnis der in fremdsprachlichen Texten verwendeten Lexeme enorm steigen kann.⁹¹³

⁹¹² Vgl. Buckingham, Louisa (2009): *Las construcciones con verbo soporte en un corpus de especialidad*. Frankfurt am Main: Lang.

⁹¹³ Vgl. Davies (2005): op. cit., S. 106.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

Davies führte eine ähnliche Studie mit spanischem Vokabular durch. Er erstellte mithilfe des CdE:Orig (siehe Kapitel 2.1.1.3.) ein Frequenzwörterbuch und verglich die im Korpus enthaltenen Texte mit dem Vorkommen der häufigsten Lexeme des Korpus. Für die Auswahl des Vokabulars wurden neben dem jeweiligen Register (gesprochene, fiktive und nicht-fiktive Texte) auch die Abdeckung der unterschiedlichen Wortarten, das Textgenre sowie die Bandbreite des Vokabulars (Vorkommen in mindestens 60 der 100 gleich großen Textblöcke des Korpus) beachtet.⁹¹⁴

Das auf diese Weise entstandene *Routledge Frequency Dictionary of Spanish* ist unterteilt in drei nach verschiedenen Kriterien eingeteilte Indizes und 30 thematische Vokabellisten. Der größte Teil des Wörterbuchs ist dem Häufigkeitsindex gewidmet, in welchem auf 170 Seiten die häufigsten spanischen Lexeme nach deren Frequenz aufgelistet sind. Der zweite Teil ist der alphabetische Index zur Erleichterung der Suche nach bestimmten Begriffen und der dritte Index ist nach Wortarten unterteilt. Die Aufteilung des Frequenzwörterbuchs in drei verschiedene Indizes erscheint sinnvoll, da die Häufigkeit des Vorkommens eines Lexems laut Mark Davies zwar eine gute Richtlinie für das Vokabellernen darstellt, aber eben nicht die einzige sein sollte. Er empfiehlt den kombinierten Gebrauch der Indizes. Wenn ein Lerner ein Pronomen im Häufigkeitsindex findet, kann er im Wortartenindex nach weiteren Pronomen und deren Häufigkeit suchen und sich so ein ganzheitliches Bild der Landschaft der spanischen Pronomina verschaffen.⁹¹⁵

Der Hauptindex des Wörterbuchs enthält Einträge, die nach Häufigkeit gestaffelt sind. Die folgende Tabelle zeigt einen Ausschnitt aus dem *Routledge Frequency Dictionary of Spanish*, der die Lexeme von Listenplatz 2000 bis 2019 zeigt:

2000 índice <i>nm</i> index, forefinger • un país pequeño, pobre, con un índice de desarrollo humano bajísimo 91 746	2010 absurdo <i>adj</i> absurd • pertenecen a otra época: extraña, absurda y romántica 90 668
2001 salvaje <i>adj</i> wild, savage, uncultivated • esos son salvajes y esos son animales domésticos 89 633	2011 dramático <i>adj</i> dramatic • el final de la comedia fue dramático 91 672

⁹¹⁴ Vgl. Davies (2005): op. cit., S. 108-113.

⁹¹⁵ Vgl. Davies (2006): op. cit., S. vii-viii.

<p>2002 marzo <i>nm</i> March</p> <ul style="list-style-type: none"> • esto ocurrió desde principios de marzo o un poco después 90 1088 	<p>2012 palo <i>nm</i> stick, pole</p> <ul style="list-style-type: none"> • tres palos de duraluminio de 32,5 metros de longitud 88 756
<p>2003 atmósfera <i>nf</i> atmosphere, environment</p> <ul style="list-style-type: none"> • tiene la atmósfera mística de un templo 88 1005 	<p>2013 diseño <i>nm</i> design</p> <ul style="list-style-type: none"> • se emplearon diseños geométricos en blanco y negro 91 945 -f
<p>2004 planeta <i>nm</i> planet</p> <ul style="list-style-type: none"> • Marte es el planeta más próximo con una superficie visible 87 1152 	<p>2014 liberal <i>adj</i> liberal</p> <ul style="list-style-type: none"> • fue más salvaje que el liberal o comunista 85 1412
<p>2005 industrial <i>adj</i> industrial</p> <ul style="list-style-type: none"> • es siempre una economía industrial o comercial 87 1829 +o -f 	<p>2015 lógica <i>nf</i> logic</p> <ul style="list-style-type: none"> • mi manera de ser tiende a la razón, a la lógica, al pensamiento 90 661
<p>2006 orgullo <i>nm</i> pride</p> <ul style="list-style-type: none"> • siente el orgullo de ser hija de un pueblo inmortal 91 738 	<p>2016 herencia <i>nf</i> inheritance, legacy</p> <ul style="list-style-type: none"> • ese carácter era una herencia de su bisabuela nórdica 92 583
<p>2007 jornada <i>nf</i> (working) day, shift</p> <ul style="list-style-type: none"> • entró en casa después de una jornada laboral especialmente agotadora 86 730 	<p>2017 legal <i>adj</i> legal</p> <ul style="list-style-type: none"> • al pueblo una norma legal para empresarios y obreros 91 987
<p>2008 desnudo <i>adj</i> nude, naked</p> <ul style="list-style-type: none"> • dibujaba la silueta del cuerpo desnudo 86 1316 	<p>2018 artificial <i>adj</i> artificial</p> <ul style="list-style-type: none"> • los fotógrafos usan luz artificial para iluminar las escenas 92 654
<p>2009 querido <i>adj</i> dear, beloved</p> <ul style="list-style-type: none"> • nuestra querida patria, la hermosa isla de Cuba 85 777 -nf 	<p>2019 deporte <i>nm</i> sport</p> <ul style="list-style-type: none"> • el béisbol, ya un deporte olímpico, se juega en muchos países 87 1141 +o

Abbildung 77: Die Rangplätze 2000 bis 2019 im *Routledge Frequency Dictionary of Spanish*.⁹¹⁶

Ein Wörterbucheintrag beinhaltet den Rang im 20 Mio. Lexeme umfassenden Korpus (beispielsweise 2019 beim letzten Eintrag der obigen Abbildung), die Wortart (*nm* für ein maskulines Substantiv), die englische Übersetzung (,sport'), ein Beispiel aus dem 100 Mio. Lexeme umfassenden CdE:Orig (Gesamt) (*el béisbol, ya un deporte*

⁹¹⁶ Tabelle: Siehe Davies (2006): op. cit., S. 78-79.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

olímpico, se juega en muchos países), die Anzahl der 100 gleich großen Wortblöcke aus dem Korpus, in denen das Lexem vorkommt (87), die Häufigkeit im Korpus (1141) und, sofern zutreffend, die Angabe des Registers, in welchem der Begriff sehr häufig oder selten auftritt, falls ein signifikanter Unterschied zu den anderen Registern besteht (+o, also häufig in mündlichen Texten).⁹¹⁷ Die zehn häufigsten Lexeme im Korpus sind im Übrigen *el/la, de, que, y, a, en, un, ser, se* und *no*.⁹¹⁸

3.3.2.2. Entwicklung von Lehrplänen und Unterrichtsmaterialien

Die Nutzung eines Korpus als Quelle für Lehrmaterialien oder für die Gestaltung des Lehrplans wird auch als COBUILD-Ansatz bezeichnet.⁹¹⁹ Guy Aston begründet den Gebrauch von korpusbasierten Unterrichtsmaterialien im Englischunterricht mit den Erkenntnissen aus Analysen großer Korpora der zeitgenössischen englischen Sprache, welche wiederholt gezeigt hätten, dass traditionelle Lehrpläne und Materialien viele linguistische Eigenschaften ignorierten, welche in muttersprachlichem Sprachgebrauch häufig auftreten würden, und andere, relativ seltene Charakteristika betonten. Seiner Meinung nach ist die Häufigkeit zwar kein so wichtiges Auswahlkriterium wie die Produktivität, dennoch ist sie nicht zu verachten, vor allem in Bezug auf das Training rezeptiver Fähigkeiten beim Lesen und Zuhören. Außerdem gibt es in den Lehrmaterialien laut Aston einige sprachliche Muster, welche die sprachliche Realität inadäquat darstellen oder ihr widersprechen.⁹²⁰

Edelhoff (1985) kritisierte bereits herkömmliche Lehrwerke, diese Kritik ist jedoch teilweise noch heute aktuell:

„Häufig werden ja in Lehrmaterialien Texte angeboten, die nur in Lehrwerken vorkommen und an denen bestimmte lexische, prosodische und grammatische Erscheinungen geübt werden, ohne daß [sic!] sie das Verstehen und Bearbeiten von Texten üben. Sie sind vielmehr Vehikel für einzelne Lehr-Items, so daß [sic!] das Wiedererkennen von Textsorten im späteren Leben und die Anwendung etwa aus der Schule bekannter Texterschließungs- und Verarbeitungstechniken eher zufällig ist.“⁹²¹

⁹¹⁷ Vgl. Davies (2005): op. cit., S. 113.

⁹¹⁸ Vgl. Davies (2006): op. cit., S. 12.

⁹¹⁹ Vgl. Ciesielska-Ciupek (2003): op. cit., S. 522-523.

⁹²⁰ Vgl. Aston (2000): op. cit., S. 8.

⁹²¹ Edelhoff (1985): op. cit., S. 7.

Weiter führt er aus: „Umgang mit Texten ist aber – wie in der eigenen Sprache – eine wesentliche kommunikative Tätigkeit in der Fremdsprache, deren Beherrschung man nicht dem ‚Leben‘ überlassen, sondern von Anbeginn des Fremdsprachenunterrichts einüben sollte.“⁹²² Er geht dabei von einem sehr weit gefassten Textbegriff aus, der Druck-, Hör- und Hör-Seh-Textsorten umfasst, und führt weiter aus, dass der Text, mangels eines unmittelbaren Zugangs zur Fremdsprache, so bedeutsam ist, weil er dadurch das eigentliche Vehikel des Fremdsprachenlernens darstellt.⁹²³

Dass die obige Kritik Edelhoffs an den Unterrichtsmaterialien zumindest in Teilen immer noch aktuell ist, kann an De Florio-Hansens Aufzählung der Gründe für die Künstlichkeit von Lehrwerkdialogen gesehen werden. Sie gibt vier Hauptgründe für diese Künstlichkeit an: Zunächst müssen die Grammatikeinheiten in Texte ‚gekleidet‘ werden. Weiter sollen die Vokabeln und Strukturen nach einer formallinguistischen Systematik eingeführt werden, wie beispielsweise das Erlernen regelmäßiger vor unregelmäßigen Verben trotz des weitaus häufigeren Vorkommens der letzteren. Außerdem sollen lexikogrammatische Einheiten in den kurzen Texten besonders häufig vorkommen, um leichter erkannt zu werden und deren Einprägung zu fördern. Schließlich sollen die didaktisierten Texte auch relativ einfach sein, um die Schüler nicht zu überfordern.⁹²⁴ Rössler geht in ihrem Aufsatz „Zur Frage der Authentizität von Lehrwerkdialogen für den Französischunterricht“ genauer auf die Authentizität nächstsprachlicher Dialoge in Lehrbüchern ein und kommt ebenfalls zu dem wenig überraschenden Schluss, dass diese stark vereinfacht sind und nur begrenzt Charakteristika gesprochener fremdsprachlicher Texte aufweisen.⁹²⁵ Daran kann man erkennen, dass auch heute Lehrwerkdialoge eigens für Unterrichtsmaterialien erstellt und keine realen Texte verwendet werden, was die Dialoge weniger künstlich erscheinen ließe. Padrós/Biechele begründen die Verwendung realer Texte mit der Vermittlung landeskundlicher Authentizität (siehe Kapitel 3.2.1.). Solche Texte, welche in der Alltagsrealität des Zielsprachenlands tatsächlich vorkämen, ermöglichten eine genaue Wiedergabe der Sprachwirklichkeit, im Gegensatz zu

⁹²² Edelhoff (1985): op. cit., S. 7.

⁹²³ Vgl. Edelhoff (1985): op. cit., S. 8.

⁹²⁴ Vgl. De Florio-Hansen, Inez (2010): „Authentizität und Kompetenzorientierung beim Lernen und Lehren von Fremdsprachen“. In: Frings, M./ Leitzke-Ungerer, E. (Hrsg.) (2010): *Authentizität im Unterricht romanischer Sprachen*. Stuttgart: ibidem-Verlag, S. 268.

⁹²⁵ Vgl. Rössler (2010): op. cit., S. 27-43.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

denjenigen, die nur zum Zweck des Fremdsprachenunterrichts erstellt würden.⁹²⁶ Zudem gibt De Florio-Hansen an, Schüler seien eher bereit einen höheren Schwierigkeitsgrad bei realen Texten zu akzeptieren, wenn das Thema derselben sie interessiere und ihnen nützlich erscheine. Weiter werden sie von realen Materialien im Gegensatz zu didaktisierten durch die Tatsache motiviert, dass sie Texte verstehen können, die von Muttersprachlern verfasst wurden. Außerdem weist De Florio-Hansen darauf hin, dass der höhere Schwierigkeitsgrad realer Texte auch lernförderlich sein kann, da einige Schüler von den üblichen Lehrwerktexten unterfordert sind.⁹²⁷

Leitzke-Ungerer legt dar, dass didaktische oder speziell für den Unterricht geschriebene Texte zwar häufig von Muttersprachlern verfasst werden, meist aber einen geringeren Grad an Authentizität aufweisen, da sie sich vor allem in den ersten Lernjahren weniger an der zielkulturellen Realität als an den eingeschränkten sprachlichen und kulturellen Kompetenzen der Schüler orientieren. Didaktisierte Texte wirken daher häufig künstlich und konstruiert, weshalb Leitzke-Ungerer zur Verwendung adaptierter Originaltexte rät. Allerdings ist sie auch der Meinung, dass fortgeschrittene Lerner durchaus mit didaktisierten Texten, die ihrem Lernkontext angemessen sind, arbeiten können, sofern diese in hohem Grad mit zielsprachlichen Normen übereinstimmen.⁹²⁸

Da die Schüler im Fremdsprachenunterricht auf Realsituationen, auf den Kontakt mit Muttersprachlern und das Lesen muttersprachlichen Materials, wie beispielsweise Zeitungen, vorbereitet werden sollen, ist es folglich ratsam im Unterricht mit vielen realen Materialien zu arbeiten. Daher erscheint die Arbeit mit Texten aus Korpora sowie die Suche in Korpora generell als gute Vorbereitung auf die Begegnung mit Muttersprachlern und muttersprachlichen Materialien.

Heloisa Collins schreibt in ihrem Aufsatz "Materials design and language corpora", dass es zwei Arten von Beiträgen der Korpuslinguistik und der TALC-Forschung (*Teaching and language corpora*) für den Sprachunterricht gibt. Einerseits stellen sie realistische Beispiele des Sprachgebrauchs zur Verfügung. Für den Englischunterricht bestätigt sie, dass Textbücher, welche nur erfundene Beispiele enthalten und deren Beschreibungen nur auf Intuition oder Material aus zweiter Hand beruhen, nach und nach durch empirische Materialien ersetzt werden, welche sich auf Korpora und andere

⁹²⁶ Vgl. Padrós/ Biechele (2003): op. cit., S. 42.

⁹²⁷ Vgl. De Florio-Hansen (2010): op. cit., S. 273.

⁹²⁸ Vgl. Leitzke-Ungerer (2010): op. cit., S. 14.

Quellen realer Sprachdaten stützen, um Beispiele und Erklärungen zu formulieren. Andererseits ermöglicht uns die Korpuslinguistik eine kritische Analyse der Lehrmaterialien, welche über Belange der Lehrmethode hinausgeht, und auch eine Bewertung dessen miteinschließt, was gelehrt wird und wie es erklärt wird. Es werden diejenigen Ausschnitte aus der Fremdsprache, die gelehrt werden, mit deren Häufigkeit in den Korpora verglichen, und aus dem Vergleich kann darauf geschlossen werden, ob die bisher gelernten Sprachausschnitte auch wirklich lernenswert sind.⁹²⁹ Klöden gibt in diesem Zusammenhang an, es sei in erster Linie der romanistischen Korpuslinguistik zu verdanken, dass die gesprochene Sprache eine Aufwertung erfuhr und inzwischen im Vergleich zur Schriftsprache als gleichwertig angesehen wird. Dieser Wechsel brachte nicht nur ein Umdenken in der Sprachwissenschaft, sondern auch im Sprachunterricht sowie in der Sprachdidaktik mit sich und er führte vom damals vorherrschenden normativen Standpunkt zu einer eher deskriptiven Sichtweise.⁹³⁰

„Häufig“ bedeutet im Kontext der Entwicklung von Lehrplänen und Unterrichtsmaterialien nicht automatisch „pädagogisch nützlich“, da für Entscheidungen im Bereich des Unterrichts unter anderem allgemeine Unterrichtsziele, die konkrete Situation der Lerner, die Lernbarkeit und die Intuition des Lehrers zu berücksichtigen sind.⁹³¹ Dennoch können Häufigkeitsanalysen aus Korpora dabei helfen, zu entscheiden, was zu welchem Zeitpunkt gelernt werden sollte.

3.3.2.2.1. Aktuelle Lehrpläne und Unterrichtsmaterialien

Im Bereich der Entwicklung von Lehrplänen und Unterrichtsmaterialien für Fremdsprachen gibt es für den englischsprachigen Bereich bereits einige Studien. Mindt fand beispielsweise heraus, dass an deutschen Schulen eine Art Schulenglisch gelehrt wird, das außerhalb des Klassenzimmers nicht existiert. Er schreibt, die Schulbücher des Fachs Englisch für deutsche Schulen beruhten auf einem strukturellen Lehrplan, welcher nicht öffentlich diskutiert werde. Obwohl es keine allgemein akzeptierten Kriterien für die Auswahl und Bewertung von Grammatikeinheiten im Lehrplan gebe, seien die

⁹²⁹ Vgl. Collins, Heloisa (2000): „Materials design and language corpora: a report in the context of Distance Education“. In Burnard, L./ McEnery, T. (Hrsg.) (2000): *Rethinking language pedagogy from a corpus perspective: papers from the Third International Conference on Teaching and Language Corpora*. Frankfurt am Main: Peter Lang, S. 52.

⁹³⁰ Vgl. Klöden (2002): op. cit., S. 9.

⁹³¹ Vgl. Kaltenböck/ Mehlmauer-Larcher (2005): op. cit., S. 78.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

Lehrpläne aufgrund der Tradition, einer stillschweigenden Übereinkunft der Lehrbuchautoren oder der Prognosen der Autoren über die jeweilige Bewertung der Fremdsprachenlehrer oft auffällig ähnlich. Das oberste Ziel des Englischunterrichts an deutschen Schulen ist die Kommunikation in Alltagssituationen. Daher sollte der Lehrplan vorwiegend auf gesprochener Sprache basieren und nur sekundär auf geschriebenem Englisch. Mindt schlägt zudem vor, dass die Progression des Lehrplans sich an Häufigkeitsanalysen aus Korpora orientieren sollte.⁹³² Auch Rüschhoff/Wolff (1999), Römer (2004), Alonso Pérez-Ávila (2007), Siepmann (2009) und Braun/Chambers (2009) kritisieren die Diskrepanzen zwischen realen Sprachen und deren Darstellung in Schulbüchern und sprechen sich für die Nutzung von Korpusdaten zur Erstellung von Lehrwerken aus.⁹³³ Des Weiteren gibt es für die englische Sprache einige korpusbasierte Studien zum Unterschied zwischen dem realen Sprachgebrauch und der Lehrwerkssprache in verschiedenen, vorwiegend die Grammatik betreffenden Bereichen, beispielsweise bei Tesch (1990), Haase (1995), Müller (2005) und Römer (2005).⁹³⁴ Diesbezüglich gibt Tolchinsky an, dass die Unterschiede zwischen dem Inhalt der Lehrbücher und den Daten des Korpus herangezogen werden können, um die Kommunikationsschwierigkeiten zwischen Muttersprachlern und Fremdsprachenlernern zu erklären. Im Vergleich zu Lehrbüchern stellen Korpora einen realistischeren Sprachgebrauch dar, der die Komplexität und die unterschiedlichen Nuancen natürlicher Sprache reflektiert.⁹³⁵

Auch De Florio-Hansen bemängelt die Unterrepräsentation bestimmter, in Korpora realer Sprache nachgewiesener Formen in Lehrwerken zugunsten der Intuition der Autoren. Sie führt aus, informelle Gespräche, welche in realen Kommunikationssituationen eine wichtige Rolle spielten und in Korpora gut belegt seien,

⁹³² Vgl. Mindt, Dieter (1996): „English corpus linguistics and the foreign language teaching syllabus“. In Thomas, J./ Short, M. (Hrsg.) (1996): *Using Corpora for Language Research*. London: Longman, S. 232-233, 245-246.

⁹³³ Vgl. Rüschhoff/ Wolff (1999): op. cit., S. 169. / Römer, Ute (2004): „Comparing real and ideal language learner input: The use of an EFL textbook corpus in corpus linguistics and language teaching“. In: Aston, G./ Bernardini, S./ Stewart, D. (Hrsg.) (2004): *Corpora and Language Learners*. Amsterdam [u.a.]: John Benjamins, S. 151-168. / Alonso Pérez-Ávila (2007): op. cit. (06.06.2017), S. 6-7. / Siepmann (2009): op. cit., S. 326. / Braun/ Chambers (2009): op. cit., S. 330.

⁹³⁴ Vgl. Tesch, Felicitas (1990): *Die Indefinitpronomina „some“ und „any“ im authentischen englischen Sprachgebrauch und in Lehrwerken*. Tübingen: Narr.

⁹³⁵ Vgl. Tolchinsky (2014): op. cit., S. 13-14.

stunden im Gegensatz zur „sicheren, harmonischen, ungestörten sprachlichen Welt der Lehrwerke“. ⁹³⁶ Ihrer Meinung nach ist die Überzeugung ‚Vom Einfachen zum Schwierigen‘ in der Pädagogik und in der Fremdsprachendidaktik tief verwurzelt und daher werden reale Texte wohl in näherer Zukunft nicht so schnell ganzheitlich in Lehrwerke integriert werden. Für sie sind vor allem die für Gespräche charakteristischen Frageformen und Wiederholungen in Lehrbüchern unterrepräsentiert. Diese sind allerdings besonders bedeutend für den Aufbau und die Pflege zwischenmenschlicher Beziehungen. Außerdem würde in didaktisierten Texten ein falsches Bild der Zielkultur vermittelt. In Lehrwerken wird die Zielkultur so dargestellt, wie sie sein sollte, weshalb die eigentlichen Probleme der Gesellschaft, welche sich auch in der Lexik und der Grammatik niederschlagen, in den Texten ausgeschlossen werden. ⁹³⁷ Ein weiterer Kritikpunkt ist der Mangel an geeigneten Konzepten zum Ausdruck von Aspekten wie ‚Einladungen verstehen‘, ‚Zustimmung ausdrücken‘ oder ‚sich beschweren‘. Häufig fehlt es an der Übermittlung verschiedener Nuancen oder des Unterschieds zwischen direkter und indirekter Vermittlung der eigenen Haltung. Des Weiteren würden dargestellte Diskursmuster derart reduziert, dass die Einübung derselben den Schülern bei ähnlichen Situationen im fremdsprachigen Ausland nur begrenzt weiterhelfen könne. ⁹³⁸

Grundsätzlich ist das im Unterricht gesprochene Spanisch ebenfalls relativ unabhängig von demjenigen, welches in Spanien gesprochen wird. Das von den Schülern gesprochene und geschriebene Spanisch beruht vollkommen auf dem Schulbuch und dem sprachlichen Modell der Lehrkraft. Da nun das Lehrbuch einen beträchtlichen Einfluss auf die Sprachentwicklung der Schüler hat, sollte dieses doch zumindest auf sprachlichen Tatsachen beruhen und nicht auf einer Art sprachlicher Schulbuchtradition, wie Mindt dies für den Englischunterricht an deutschen Schulen nachgewiesen hat. Aus diesem Grund sollten die in Schulbüchern und Unterrichtsmaterialien enthaltenen Sprachausschnitte zumindest größtenteils aus realen Materialien stammen, welche eben aus Korpora entnommen und bearbeitet werden können. Sofern die im Unterricht verwendeten Materialien nicht der aktuell im spanischsprachigen Ausland gesprochenen Sprache entsprechen, werden die Schüler beim direkten Kontakt mit Muttersprachlern diese Unterschiede feststellen und manche gelernten Inhalte erst einmal an den aktuellen

⁹³⁶ De Florio-Hansen (2010): op. cit., S. 268.

⁹³⁷ Vgl. De Florio-Hansen (2010): op. cit., S. 268-269.

⁹³⁸ Vgl. De Florio-Hansen (2010): op. cit., S. 270-272.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

Sprachstand anpassen müssen. So kann ein Schüleraustausch oder der Kontakt mit Spaniern im Urlaub relativ ernüchternd, demotivierend und enttäuschend sein.

Mindt führte außerdem eine Untersuchung zu den *irregular verbs* des Englischen durch und kam zu dem Ergebnis, dass Lerner des Englischen allein mit den zehn häufigsten unregelmäßigen Verben (*say, make, go, take, come, see, know, get, give* und *find*) mehr als 45%, also fast die Hälfte aller Formen der unregelmäßigen Verben des Englischen beherrschten.⁹³⁹ Dies bedeutet, dass mit einem relativ geringen Lernaufwand ein verhältnismäßig großer Teil der Verben in englischen Texten verstanden werden kann, ähnlich der Frequenzanalysen zum spanischen Vokabular und der Erstellung des *Routledge Frequency Dictionary of Spanish* (siehe Kapitel 3.3.2.1.).

Genau aus diesem Grund könnten auch die bisher im Spanischunterricht an deutschen Schulen verwendeten Lehrwerke auf deren Übereinstimmung mit dem Sprachgebrauch spanischer Muttersprachler verglichen werden, um herauszufinden, ob die von Lehrbuchautoren als wichtig empfundenen sprachlichen Strukturen und Lexeme auch wirklich im Umgang mit spanischen Medien oder für die Kommunikation mit Spaniern und Lateinamerikanern relevant sind. Auch wenn die Schulbuchtradition nicht sofort komplett geändert werden kann und sollte, lockert der gelegentliche Einsatz von Arbeitsblättern mit realen Beispielen aus Korpora den Unterricht auf und holt die fremdsprachliche Welt ins Klassenzimmer.

Ogleich die Frequenz alleine nicht das ausschlaggebende Kriterium für die Erstellung von Lehrplänen sein kann, so ist auch Hunston der Meinung, dass das uns aus Korpora bekannte Wissen zu einer Veränderung der Lehrplanerstellung führen sollte.⁹⁴⁰ Sie schlägt den sogenannten *lexical syllabus* von Sinclair/Renouf (1988)⁹⁴¹ beziehungsweise Willis (1990)⁹⁴² vor. Willis/Willis haben außerdem ein auf diesem lexikalischen Lehrplan beruhendes Lehrbuch veröffentlicht, den *Collins COBUILD English Course*.⁹⁴³ Es handelt sich dabei um einen lexikalischen Lehrplan, der nicht nur

⁹³⁹ Vgl. Mindt (1996): op. cit., S. 233.

⁹⁴⁰ Vgl. Hunston (2002): op. cit., S. 189.

⁹⁴¹ Vgl. Sinclair, John McHardy/ Renouf, Antoinette (1988): „A lexical syllabus for language learning“. In: Carter, R. / McCarthy, M. (Hrsg.) (1988): *Vocabulary and language teaching*. Harlow: Longman, S. 140–158.

⁹⁴² Vgl. Willis, Dave (1990): *The lexical syllabus: A new approach to language teaching*. London: HarperCollins.

⁹⁴³ Vgl. Willis, Jane/ Willis, Dave (1989): *Collins COBUILD English Course*. London: HarperCollins.

aus Vokabular besteht, wie missverständlicherweise angenommen werden könnte, sondern alle Bereiche der Sprache abdeckt. Die Lexik ist lediglich das zentrale Organisationskonzept. Der Gedanke hinter diesem Ansatz ist, dass die häufigsten Lexeme einer Sprache zuerst unterrichtet werden sollten.⁹⁴⁴ Daher schlagen Sinclair/Renouf vor, dass das Hauptaugenmerk auf a) die am weitesten verbreiteten Wortformen einer Sprache, b) die zentralen Gebrauchsmuster und c) die Kombinationen, welche sie gewöhnlich bilden, gelegt werden sollte.⁹⁴⁵ Nach Hunstons Meinung wären mit den frequentesten Lexemen viele unterschiedliche Kombinationen möglich, wodurch die Lerner sehr schnell eine gewisse Flexibilität erlangen könnten. Zusätzlich würden diese Hauptverwendungsweisen der häufigsten Begriffe gleichzeitig die wichtigsten Grammatikeinheiten abdecken, wenn auch auf ungewöhnliche Art und Weise. Das Verb *make* hat beispielsweise viele Verwendungsweisen, von denen einige nur selten in Anfangskursen vermittelt werden. So finden sich häufiger Kombinationen wie *make decisions*, *make discoveries*, *make arrangements* als die konkrete Bedeutung des Verbs in *make a cake*. *Make* ist im ersten Fall ein *delexical verb* und wird in dieser Bedeutung häufiger verwendet als im Sinne des gewöhnlichen Verbs mit der Bedeutung ‘machen, tun’. Wenn im Unterricht lediglich die konkrete Bedeutung von *make* vermittelt wird, bekommen die Lerner keine Gelegenheit dazu, differenziertere Bedeutungen mit einem einfachen Verb auszudrücken.⁹⁴⁶

Auch in der spanischen Sprache gibt es derartige Verben mit einer konkreten Bedeutung einerseits und unterschiedlichen Bedeutungen als Bestandteil mehrgliedriger Ausdrücke andererseits. Ein Beispiel hierfür ist das spanische Pendant zum englischen *make*, das Verb *hacer*. Neben der Bedeutung ‘tun, machen’ ist es auch Teil der Kombinationen *hacer los platos* (‘Geschirr spülen’), *hacer puerto* (‘einlaufen’), *hacer bien/mal* (‘guttun/schlecht tun’) oder *hacerse* (‘werden’). Ein anderes Beispiel ist das Verb *poner*, welches neben der gewöhnlichen Bedeutung ‘setzen, stellen, legen’ auch in Kombinationen wie *poner la mesa* (‘den Tisch decken’), *poner la radio* (‘das Radio anschalten’) oder *poner buena/mala cara* (‘gute/böse Miene machen’) vorkommt. Es ist also ebenfalls bei einigen Verben des Spanischen von Vorteil, nicht nur die konkrete

⁹⁴⁴ Vgl. Hunston (2002): op. cit., S. 189.

⁹⁴⁵ Vgl. Sinclair/ Renouf (1988): op. cit., S. 148.

⁹⁴⁶ Vgl. Hunston (2002): op. cit., S. 189.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

Bedeutung, sondern auch die häufigsten Kombinationen, in denen das Lexem vorkommt, zu erlernen.

Siepmann sieht das ähnlich. Er schreibt, die Korpuslinguistik beweise, dass der Sprachbenutzer nicht beliebig viele Sätze unter Einhaltung einiger weniger syntaktisch-semanticischer Regeln hervorbringen kann, wie Chomsky es angenommen hat, sondern dass er lediglich über eine sehr eingeschränkte Freiheit im Sprachgebrauch verfügt. Statt einer Grammatik, welche die Gesamtheit der Regeln in sich vereint, und einem Lexikon, welches aus Lexemen und Wendungen aufgebaut ist, sei eine Lexikogrammatik anzunehmen, welche aus typischen grammatischen Verbindungen, sogenannten *Kolligationen* (z. B. *noun + is about + verb-ing: probability is about calculating the likelihood of various outcomes from particular events*), und typischen lexikalischen Verbindungen, sogenannten *Kollokationen* (z. B. *clear motorway, richly embroidered, make babies, I couldn't agree more*), bestehe (siehe auch Kapitel 4.1.1.). Es stellt sich nicht die Frage nach der Grammatikalität oder Ungrammatikalität von Äußerungen, sondern danach, ob die sprachlichen Elemente eine Kollokation oder Kolligation bilden. Zudem besteht eine enge Verbindung von Semantik und Syntax, wobei die Wahl der Syntax nicht frei ist, sondern von der Lexik bestimmt wird.⁹⁴⁷ „Die Korpuslinguistik verweist damit auf die Notwendigkeit, den Wortschatz in das Zentrum aller fremdsprachendidaktischen Bemühungen zu stellen und, wo immer möglich, Wortschatz und Grammatik unterrichtlich zu vereinen“, so Siepmann.⁹⁴⁸

Ein Fokus des lexikalischen Ansatzes liegt auf der Vermittlung von eben solchen Kollokationen und dem damit verbundenen Konzept ‚vorgefertigter Einheiten‘. Es besteht Einigkeit darüber, dass das Wissen über Kollokationen wichtig ist für die Entwicklung sprachlicher Fähigkeiten, sowohl in der Muttersprache als auch in der Fremdsprache. Das Wissen über Kollokationen gibt an, welche lexikalischen Einheiten häufig mit anderen vorkommen und wie sie sich in einem Satz verbinden.⁹⁴⁹ Daher benötigen Lerner detaillierte Informationen über gängige Kollokationen und Phraseologismen, verschiedene Grade der Variabilität, die relative Häufigkeit und Geläufigkeit bestimmter Muster sowie den Grad der Förmlichkeit.⁹⁵⁰ In dieser Hinsicht

⁹⁴⁷ Vgl. Siepmann (2009): op. cit., S. 325.

⁹⁴⁸ Siepmann (2009): op. cit., S. 325.

⁹⁴⁹ Vgl. McEnery/ Xiao (2010): op. cit., S. 368.

⁹⁵⁰ Vgl. McAlpine, Janice/ Myles, Johanne (2003): „Capturing phraseology in an online dictionary for advanced users of English as a second language: A response to user needs“. In: *System*, 31, S. 75.

sind Korpora sehr nützlich, nicht nur, weil mit ihrer Hilfe quantitative Aussagen zu Kollokationen gemacht werden können, sondern auch, weil die Darstellungsweise als *Key Word in Context* (siehe Kapitel 1.4.3.) den Lernern eine große Anzahl an realen Sprachbeispielen in strukturierter Weise bietet.⁹⁵¹ Auch Tolchinsky weist auf die zunehmende Bedeutung linguistischer Korpora in der Vermittlung von Kollokationen und Phraseologismen hin.⁹⁵² Durch die Korpusstudien haben eben diese Konzepte enorm an Bedeutung gewonnen und sollten daher nun verstärkt in den Unterricht integriert werden.

Für den lexikalischen Ansatz ist es also einerseits von Bedeutung, dass das Vokabular im Kontext gelernt wird, aber andererseits auch, dass es die Interessen und Bedürfnisse der Lerner anspricht, ihre Vorkenntnisse berücksichtigt und, dass es neue Informationen bereithält. Der Unterricht sollte laut Lahuerta/Pujol als ein Prozess gesehen werden, der nicht nur die Produkte an sich, sondern auch die Vorgehensweise zeigt, wie diese Produkte erarbeitet werden. Sie sind der Meinung, es ist wichtig, dass im Unterricht nicht nur Begriffe erklärt werden, sondern auch veranschaulicht wird, wie Mitteilungen trotz unbekannter Lexeme verstanden und wie Bedeutungen erschlossen werden können. Die Schüler sollten nicht nur Sachverhalte auswendig lernen, sondern es soll ihnen gezeigt werden, wie sie diese Aufgabe am einfachsten bewerkstelligen. Zudem wirkt es sich vorteilhaft aus, wenn Sprachproduktionen nicht nur aufgegeben, sondern stimuliert und anregende Situationen zum sprachlichen Austausch geschaffen werden.⁹⁵³ Aus diesen Gründen sei das Lernen von Vokabellisten mit vielen Nachteilen verknüpft, so die Autoren.⁹⁵⁴ In diesem Zusammenhang schlägt Alonso Pérez-Ávila die Nutzung eines linguistischen Korpus zur Aneignung von Vokabular vor. In den Korpora können lexikalische Einheiten dank der Konkordanzen im Kontext gezeigt werden und sie sind zudem nach nützlichen Klassifikationen geordnet, welche es erlauben, beispielsweise verschiedene Register oder Themenbereiche zu unterscheiden. Auf diese Weise können gleichzeitig Lexik und Pragmatik gefördert werden, wodurch die Schüler erfahren, wann ein gewisser Begriff angebracht ist und wann nicht. Zudem erlauben die Korpora mittels

⁹⁵¹ Vgl. McEnery/ Xiao (2010): op. cit., S. 368.

⁹⁵² Vgl. Tolchinsky (2014): op. cit., S. 14.

⁹⁵³ Vgl. Lahuerta, Javier/ Pujol, Mercè (1996): „El lexicon mental y la enseñanza del vocabulario“. In: Segoviano, C. (Hrsg.) (1996): *La enseñanza del léxico español como lengua extranjera*. Madrid: Iberoamericana, S. 124.

⁹⁵⁴ Vgl. Lahuerta/ Pujol (1996): op. cit., S. 127.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

des Gebrauchs von Platzhaltern die Suche nach einem Lemma in all seinen morphologischen und syntaktischen Kombinationen, was dem Schüler dabei helfen kann, sich über das unterschiedliche syntaktische Verhalten verschiedener Lexeme und Wortarten bewusst zu werden sowie assoziative Netze zu erstellen, mithilfe derer er sich das Vokabular besser merken kann.⁹⁵⁵

Gómez Molina vertritt die Meinung, dass man sich für die Auswahl des im Spanischunterricht gelehrtten Vokabulars zwar auf die eigene Kompetenz verlassen könnte, es aber ratsam sei, ideologische, thematische, frequenzbasierte und phraseologische Wörterbücher des aktuellen Spanisch zu konsultieren, um die Fülle der Lexik kennenzulernen. Nachdem derartige Lexika zurate gezogen worden sind, können die neuen lexikalischen Einheiten dementsprechend ausgewählt und beispielsweise über Mindmaps den Schülern zugänglich gemacht werden, welche die neue Lexik daraufhin in ihr persönliches mentales Lexikon integrieren.⁹⁵⁶ Alonso Pérez-Ávila erweitert diese Nachschlagewerke um die Möglichkeit der Suche nach Konkordanzen in Korpora, welche für sie zudem einfacher und schneller scheint. Die Ergebnisse derartiger Suchen garantieren nicht nur Angaben zum Gebrauch und der Frequenz einer bestimmten lexikalischen Einheit, sondern sie werden zudem kontextualisiert und nach Genres, Zeitabschnitten und geografischen Zonen geordnet angezeigt.⁹⁵⁷ Grundsätzlich könnten uns Korpora somit helfen bei der Entscheidung darüber, was unterrichtet werden soll, also in Bezug auf den Lehrplan, und bei der Erstellung geeigneter Materialien, um diesen Lehrplan zu realisieren.

Laut der Aussagen der Redaktion Französisch sowie der Redaktion Spanisch/Italienisch/Russisch des Cornelsen Verlags, einem der größten Verlage im Bereich der Fremdsprachen, werden bei der Erstellung der Lehrwerke in diesen Sprachen bisher keine Korpora verwendet.⁹⁵⁸ Die Redaktionsleitung Französisch führt aus, dass ihr für das Französische derzeit keine aktuellen Korpora zum Sprachgebrauch bekannt seien,

⁹⁵⁵ Vgl. Alonso Pérez-Ávila (2007): op. cit. (06.06.2017), S. 7-9.

⁹⁵⁶ Vgl. Gómez Molina, José Ramón (2004): „Los contenidos léxico-semánticos“. In: Sánchez Lobato, J./ Santos Gargallo, I. (Hrsg.) (2004): *Vademécum para la formación de profesores. Enseñar español como segunda lengua (L2)/ lengua extranjera (LE)*. Alcobendas: Sociedad General Española de Librería, S. 799.

⁹⁵⁷ Vgl. Alonso Pérez-Ávila (2007): op. cit. (06.06.2017), S. 7.

⁹⁵⁸ Vgl. E-Mail-Kontakt mit Heike Malinowski, Redaktionsleiterin ‚Weitere Fremdsprachen‘ des Cornelsen Verlags, am 31.07.2017.

die Universitäten Paderborn und Osnabrück sich jedoch gerade mit dem Aufbau einer großen Korpusdatenbank beschäftigten, auf die der Cornelsen Verlag neue Lehrwerke stützen dürfe.⁹⁵⁹

Christoph Bürgel, Professor für Didaktik des Französischen und Spanischen an der Universität Paderborn, bedauert, dass es für die französische Sprache bislang kein repräsentatives, verschiedene Genres umfassendes Korpus gegeben hat, wie es für die englische Sprache der Fall ist. Daher sei bislang kein korpusinduzierter, variationistischer Zugriff auf das moderne Französisch möglich gewesen. Das aus diesem Grund von Bürgel/Siepmann erstellte *Corpus de référence du français contemporain* (kurz: CRFC) ist ein Referenzkorpus des Französischen Frankreichs und richtet sich an verschiedenste Nutzerkreise. Neben Linguisten sollen auch Fremdsprachendidaktiker das Korpus zur Neuentwicklung von korpusbasierten Nachschlagewerken und Lehrbüchern, Französischlehrer für die Unterrichtsplanung und als Korrekturhilfe, Übersetzer und technische Redakteure für gezielte Anfragen zu Formulierungsproblemen sowie Studierende zur Sensibilisierung für sprachliche Regelmäßigkeiten nutzen.⁹⁶⁰ Es handelt sich dabei um ein 310 Mio. Lexeme umfassendes Korpus, das eine große Bandbreite an verschiedenen Genres abdeckt und bei dem ebenso viele Einheiten aus geschriebenen Quellen wie aus gesprochenen oder sogenannten pseudo-gesprochenen Quellen (beispielsweise Transkripte von Fernsehsendungen) stammen. Es ist das erste französische Korpus, das zu einem wesentlichen Teil aus spontaner (ca. 30 Mio. Lexeme) und pseudo-gesprochener Sprache besteht (ca. 125 Mio. Lexeme). Derzeit ist das CRFC das größte französische Korpus, das nicht vollständig auf Internetquellen beruht, und es wird kontinuierlich aktualisiert mit jährlich mindestens 25 Mio. Lexemen gesprochener und geschriebener Sprache seit dem Jahr 2014. Dabei lehnt es sich hinsichtlich der Planung und Struktur an die großen Referenzkorpora des britischen und amerikanischen Englisch an (das *British National Corpus* und das *Corpus of Contemporary American*

⁹⁵⁹ Vgl. E-Mail-Kontakt mit der Redaktion ‚Französisch‘ des Cornelsen Verlags am 28.07.2017.

⁹⁶⁰ Vgl. Bürgel, Christoph: *Forschung*. Online Publikation: <https://kw.uni-paderborn.de/institut-fuer-romanistik/prof-dr-christoph-buergel/forschung/> (02.08.2017).

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

English), mit der Ausnahme, dass es hinsichtlich der enthaltenen Genres noch vielfältiger ist als jedes andere Korpus.⁹⁶¹

Die Ersteller betonen vor allem den großen Teil an gesprochener, informeller Sprache. Bisher musste in diesem Bereich auf interviewbasierte Texte zurückgegriffen werden, im CRFC sind nun jedoch 3000 Stunden natürlicher Sprache aus 6000 Fernsehsendungen enthalten, welche in den Jahren 2013 und 2014 gesammelt wurden.⁹⁶² Als Antwort auf die Frage, warum nun gerade Transkripte von Fernsehsendungen integriert werden, argumentieren sie mit Meißner, der angibt, dass im Fernsehen eine Art statistische Norm an Idiolekten und linguistischen Varietäten zu hören ist, welche dennoch von der breiten Masse verstanden werden kann. Somit kann die Breite linguistischer Varietät in den Medien als Richtschnur für die Zusammenstellung gesprochener Korpora und die Entwicklung von Unterrichtsmaterialien dienen. Meißner ist der Meinung, sofern gewisse Sprachformen in frankophonen Medien nicht gesendet werden können, weil sie nicht verstanden werden, sollten sie auch nicht im Unterricht vorkommen. Das von ihm so betitelte ‚Super-Register‘ Mediensprache gibt Aufschluss darüber, welche Idiolekte, Dialekte, Soziolekte und Sprechstile im Unterricht besprochen werden sollten.⁹⁶³ Zudem wurde bei der Korpuserstellung Wert darauf gelegt, dass keine Kategorie des geschriebenen Französisch eine Vorrangstellung gegenüber den anderen hat, wie dies laut den Autoren bei Zeitungstexten in anderen Korpora häufig der Fall sei.⁹⁶⁴ Das CRFC ist laut Bürgel bisher nur auf der Plattform *SketchEngine* zugänglich, soll aber bis Ende 2018 auf die Plattform *PrimeStat* transferiert werden.⁹⁶⁵

Bürgel/Siepmann erstellten das CRFC auch mit dem Augenmerk auf die Verbesserung von Unterrichtsmaterialien für den Französischunterricht. Es soll beispielsweise eine lexikographische Beschreibung des umgangssprachlichen Französisch sowie eine lexikogrammatische und gebrauchsbasierte Herangehensweise an die französische Grammatik ermöglichen. Wie die Autoren bereits bei Pilotstudien

⁹⁶¹ Vgl. Siepmann, Dirk/ Bürgel, Christoph/ Diwersy, Sascha (2017): „The Corpus de référence du français contemporain (CRFC) as the First Genre-Diverse Mega-Corpus of French“. In: *International Journal of Lexicography*, Volume 30, Issue 1, S. 70.

⁹⁶² Vgl. Siepmann/ Bürgel/ Diwersy (2017): op. cit., S. 71.

⁹⁶³ Vgl. Meißner, Franz-Joseph (2006): „Linguistische und didaktische Überlegungen zur Entwicklung von Kompetenzaufgaben im Lernbereich Mündlichkeit“. In: *Französisch heute* 3, S. 249-250.

⁹⁶⁴ Vgl. Siepmann/ Bürgel/ Diwersy (2017): op. cit., S. 74.

⁹⁶⁵ Vgl. Siepmann, Dirk/ Bürgel, Christoph/ Diwersy, Sascha (2016): „Le Corpus de référence du français contemporain (CRFC), un corpus massif du français largement diversifié par genres“. In: *Congrès Mondial de Linguistique Française – CMLF 2016*, S. 2.

herausfanden, könnten anhand des Korpus typische Prä- und Postmodifikationen sowie Valenzmuster für Vokabellisten ermittelt, lexikostatistische Studien zur Erstellung eines Kernvokabulars durchgeführt und die traditionelle französische intuitionsbasierte Lexikographie durch Korpusstudien vor allem im Hinblick auf frequente Bedeutungen und Kollokationen ergänzt werden. Außerdem führten sie bereits eine Pilotstudie zum *Subjonctif* durch, bei der sie zu dem Ergebnis kamen, dass der Gebrauch desselben in geschriebenem und gesprochenem Französisch teilweise erheblich von demjenigen abweicht, der in Lehrbüchern vermittelt wird, insbesondere im Hinblick auf die Auslöser des *Subjonctif*.⁹⁶⁶

Im Jahr 2015 nahm Siepmann eine auf dem CRFC basierende Untersuchung zur Repräsentation sprechsprachlicher Lexis in Wörterbüchern vor und kam dabei zu dem Schluss, dass sowohl die untersuchten ein- als auch die zweisprachigen Wörterbücher der französischen Sprache die Eigenschaften der sprechsprachlichen Lexik bisher nur teilweise erfassen.⁹⁶⁷ Bürgel/Siepmann veröffentlichten außerdem im selben Jahr einen phraseologischen Grundwortschatz des Französischen. Die „Gesamtliste der häufigsten Phraseme des Französischen“ ist im Internet einsehbar und eine Veröffentlichung der Frequenzlisten des Mündlichen und Schriftlichen sind geplant.⁹⁶⁸ Oliver Wicher arbeitet derzeit an einer Analyse französischer Lehrwerke im deutschsprachigen Raum. Für diesen Vergleich des realen Französisch mit dem sogenannten ‚Schulfranzösisch‘ werden das CRFC und das eigens für dieses Projekt erstellte *French Foreign Language Textbook Corpus* (FFLTC) verglichen, um Aussagen über sprachliche Eigenschaften der Lektionstexte, wie ihre Lexikogrammatik oder ihre Idiomatizität, machen zu können.⁹⁶⁹

Derzeit gibt es leider kein Projekt zur Erstellung eines derartigen spanischen Referenzkorpus. Auch Heike Malinowski, die Redaktionsleiterin des Cornelsen Verlags für die Fremdsprachen Spanisch/Italienisch/Russisch, bedauert es, keine Korpusdatenbank für die Erstellung der Spanischlehrwerke zur Verfügung zu haben.

⁹⁶⁶ Vgl. Siepmann/ Bürgel/ Diwersy (2017): op. cit., S. 79-81.

⁹⁶⁷ Vgl. Siepmann, Dirk (2015): „Dictionaries and spoken language: a corpus-based review of French dictionaries“. In: *International Journal of Lexicography*. Band 28, Ausgabe 2, 1. Juni 2015, S. 139-168.

⁹⁶⁸ Vgl. Bürgel, Christoph/ Siepmann, Dirk (2015): *Gesamtliste der häufigsten Phraseme des Französischen*. Online Publikation: https://kw.uni-paderborn.de/fileadmin/fakultaet/Institute/romanistik/Prof._Dr._Christoph_Buergel/Gesamtliste_Phraseme_des_Franzoesischen.pdf (02.08.2017).

⁹⁶⁹ Vgl. Wicher, Oliver: *Projekte. Schulfranzösisch – das Lehrbuchkorpus FFLTC*. Online Publikation: <https://kw.uni-paderborn.de/institut-fuer-romanistik/oliver-wicher/projekte/> (17.08.2017).

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

Ihrer Meinung nach könnten Korpora jedoch vor allem im Anfangsunterricht nur bedingt eingesetzt werden. Sie gibt an, dass Schulbücher, insbesondere in den ersten Lernjahren, hauptsächlich anhand von Themenbereichen wie ‚Familie‘, ‚Freunde‘, ‚Das eigene Zimmer‘ oder ‚Das eigene Haus‘ erstellt würden und das neue Vokabular, welches auch mengenmäßig in einem gewissen Rahmen bleiben sollte, daher bereits zu einem großen Teil festgelegt sei. Der Wortschatz sei lediglich ein Bereich der Schulbucheerstellung, daneben gäbe es auch Anforderungen an sprachliche Strukturen sowie Themenschwerpunkte, wodurch nicht mehr viel Spielraum bleibe. Zudem würden in den Lehrplänen der einzelnen Bundesländer teilweise sehr detaillierte Anforderungen an die Schulbuchverlage gestellt, wodurch die Auswahl an grammatischen Strukturen in einigen Bundesländern, wie beispielsweise im neuen bayerischen Lehrplan Plus (siehe Kapitel 3.1.3.), bereits festgelegt sei. Bei den fortgeschrittenen Lernern sieht sie jedoch im Bereich des kontextualisierten Lernens die Möglichkeit, im Unterricht anhand von Korpora die genaue Verwendung grammatischer Strukturen zu analysieren, um den Schülern eben diese Strukturen, welche häufig auch mit einer bestimmten Textsorte oder einem gewissen Register verknüpft sind, kontextualisiert zu vermitteln, wie es auch in den Schulbüchern dieser Jahrgangsstufen der Fall ist. Sie gibt zudem an, dass es bei der Erstellung von Lehrwerken aufgrund der zahlreichen Anforderungen, welche beachtet werden müssen, durchaus vorkommen kann, dass einzelne Lexeme, aufgrund von Veränderungen an Texten, herausfallen. Hier wäre es wünschenswert, eine Korpusdatenbank zur Verfügung zu haben, anhand derer man das Vokabular abgleichen könnte, um frequente beziehungsweise wichtige Lexeme herauszufiltern, sodass am Ende der Lehrbucheerstellung keine relevanten Begriffe fehlen.⁹⁷⁰

In Bezug auf die Frage, warum die derzeit verfügbaren spanischen Korpora nicht für die Lehrbucheerstellung genutzt werden könnten, gibt Frau Malinowski an, dass diese zu ‚sophisticated‘ seien. Im ersten Spanischlernjahr werden beispielsweise ungefähr 650 Lexeme gelernt und in den drei Lernjahren Spanisch als spätbeginnende Fremdsprache sind es insgesamt 2500 Lexeme, daher sind die meisten Korpora zu überdimensioniert. In den Lehrbüchern der ersten beiden Lernjahre finden sich im Wesentlichen mündliche Texte. Erst im dritten Lernjahr oder in der Oberstufe werden Textproduktionen wie der

⁹⁷⁰ Vgl. Telefonat mit Heike Malinowski, Redaktionsleiterin ‚Weitere Fremdsprachen‘ des Cornelsen Verlags, am 16.08.2017.

Comentario personal eingeführt, welche jedoch hauptsächlich in der Schule verfasst werden und danach nicht mehr. Zudem seien einige Lexeme bereits durch den Lehrplan festgelegt. So fordert der bayerische Lehrplan beispielsweise die Behandlung des Themas *México* im ersten Lernjahr, wodurch einige Begriffe wie Eigennamen und Indigenismen bereits festgelegt sind, obwohl diese rein der Frequenz wegen nicht in den Wortschatz aufgenommen würden. Aus diesem Grund wird bei der Schulbucherstellung die Vokabelauswahl von den Autoren übernommen. Nach Malinowskis Meinung müsste ein Korpus weniger nach Textsorten als vielmehr nach kommunikativen Lernzielen aufgebaut sein. So sollten die Texte nach Themen ausgewählt werden, also im Sinne eines semantischen Ansatzes. Beispiele für diese Lernziele sind die bereits erwähnten Themen der ersten Lernjahre ‚Ich‘, ‚Meine Eltern‘, ‚Mein Haus‘, ‚Mein Viertel‘, ‚Haustiere‘, welche die Schüler adäquat beschreiben sollen. Anhand eines derartigen Korpus könnte eine „alltagstaugliche Frequenzanalyse“ durchgeführt werden, welche für die Lehrbucherstellung nutzbar wäre.⁹⁷¹ Ein erstes, wenn auch relativ kleines Korpus, welches auf diese Weise aufgebaut wurde, ist das Korpus *Sacodeyl* (siehe Kapitel 2.4.3.1.).

Als Beispiel für die Auswahl der sprachlichen Einheiten in Lehrwerken gibt Heike Malinowski ein Grammatik und Wortschatz betreffendes Thema an, welches in den Spanischlehrwerken des Cornelsen Verlags abgeändert wurde. Traditionell wurden in den Lehrwerken die Einheiten *este*, *ese* und *aquel* vermittelt, doch Letzteres wird in der Alltagskommunikation eher selten verwendet. Die Antwort auf die Frage nach einem Satz mit *aquel* stelle sich laut Malinowski meist als relativ problematisch heraus, was daran liege, dass dieses Demonstrativpronomen hauptsächlich in der schriftlichen Sprache verwendet würde und meist eine temporale Bedeutung wie in *en aquellos días* hätte. Aus diesem Grund wurde es vom Cornelsen-Verlag aus der erwähnten Grammatik- und Wortschatzeinheit herausgelassen. Frau Malinowski gibt an, dass derartige Fragen im Verlag durch die Befragung der muttersprachlichen Autoren geklärt werden und eine Korpusanalyse hier wahrscheinlich nicht hilfreich wäre.⁹⁷² Das Urteil der Muttersprachler ist in diesem Punkt durchaus ein wichtiger Faktor, doch bei der Befragung des CORPES (siehe Kapitel 2.1.1.2.) ergibt sich ein ähnliches Bild. Hier

⁹⁷¹ Vgl. Telefonat mit Heike Malinowski, Redaktionsleiterin ‚Weitere Fremdsprachen‘ des Cornelsen Verlags, am 23.08.2017.

⁹⁷² Vgl. Telefonat mit Heike Malinowski am 23.08.2017. Op. cit.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

erscheint das Lemma *este* über 1,16 Mio. Mal im Korpus, *ese* knapp 752.000 Mal und *aquel* lediglich knapp 167.000 Mal. Zur Untermauerung der muttersprachlichen Intuition könnte ein Korpus hier also wertvolle Dienste leisten. In Bezug auf die vermutete hauptsächlich temporale Bedeutung von *aquel* kann das CORPES die Intuition der Mitarbeiter jedoch nicht bedingungslos bestärken. Es finden sich zwar einige Beispiele wie *en aquel tiempo* oder *aquella noche*, doch der Begleiter wird bereits in den ersten 40 aufgelisteten Konkordanzen auch lokal oder in Bezug auf Personen oder Dinge gebraucht, wie in *aquella casa*, *aquel hombre*, *aquel muro* oder *aquellas palabras*. Hier müsste allerdings eine eingehende Analyse durchgeführt werden, um aussagekräftige Ergebnisse präsentieren zu können. Man könnte jedoch durchaus sagen, dass die Nutzung eines Referenzkorpus wie des CORPES generell von Vorteil ist, um derartige Hypothesen zu untermauern oder zu entkräften.

3.3.2.2.2. Vergleich einer Studie zum Ausdruck der Zukünftigkeit mit aktuellen Schulgrammatiken

Im Hinblick auf die Vermittlung von Grammatikeinheiten in Unterrichtsmaterialien ist auch Sedanos bereits erwähnte Studie zu nennen. Sie untersuchte die Verwendung des morphologischen Futurs (*cantaré*) und des periphrastischen Futurs (*voy a cantar*) in Bezug auf die zeitliche Distanz, die grammatische Person in der Zukunftszeit und den angenommenen Wahrheitsgehalt der Aussage. Ihre Ergebnisse, speziell aus der Presse und gesprochenen Konversationen, zeigen große Unterschiede in der Verwendung dieser beiden Zukunftszeiten abhängig vom Gebrauchskontext und der Absicht des Schreibers oder Sprechers.⁹⁷³

Sie kommt zu dem Schluss, dass in gesprochenem Spanisch generell das periphrastische Futur frequenter ist. Außerdem wird es in Lateinamerika sehr viel häufiger gebraucht als das morphologische Futur, teilweise sogar ausschließlich. Im Gegenzug findet sie in Korpora geschriebener Sprache häufiger das morphologische Futur. Die unterschiedliche Verwendung der beiden Formen ist in literarischen Texten auch vom Bildungsgrad der Charaktere sowie der Formalität der Äußerung abhängig. So wird das morphologische Futur tendenziell von gebildeten Charakteren in formellen Situationen und das periphrastische Futur von weniger gebildeten Personen in

⁹⁷³ Vgl. Sedano (2007): op. cit., S. 132-144.

informellen Situationen bevorzugt. Zudem ist die Verwendung der Zukunftsform in geschriebenem Spanisch stark vom jeweiligen Autor abhängig.⁹⁷⁴

Bei der Untersuchung eines Korpus des gesprochenen Spanisch und eines Korpus mit Zeitungstexten, beide mit Texten aus Venezuela, kam sie ebenfalls zu dem Schluss, dass das morphologische Futur hauptsächlich in geschriebenem Spanisch und das periphrastische Futur größtenteils in der gesprochenen Sprache verwendet wird.⁹⁷⁵

Im Hinblick auf den zeitlichen Abstand zwischen der Äußerung selbst und dem Zeitpunkt, auf den sich der Sprecher oder Schreiber bezieht, kann sie feststellen, dass für die unmittelbare Zukunft (z. B. *en seguida, en un momento*) ausschließlich das periphrastische Futur verwendet wird, für die relativ nahe Zukunft (z. B. *esta tarde, mañana*) in gesprochenen Texten das periphrastische und in geschriebenen Texten das morphologische Futur bevorzugt wird und für weiter entfernte Zukünftigkeit (z. B. *algún día, tarde o temprano*) ausschließlich das morphologische Futur im geschriebenen Korpus und das periphrastische Futur häufiger als das morphologische Futur im gesprochenen Korpus vorkommt.⁹⁷⁶

Bezüglich der Abhängigkeit von der grammatischen Person findet sie heraus, dass die erste Person mit größerer Sicherheit über ihre Absichten spricht, während dies bei der zweiten und dritten Person nicht möglich ist, was den hauptsächlichsten Gebrauch des periphrastischen Futurs in der ersten Person erklärt.⁹⁷⁷

Hinsichtlich des Wahrheitsgehalts kommt Sedano zu dem Ergebnis, dass das morphologische Futur vor allem bei Unsicherheit bezüglich der Aussage verwendet wird, während bei Sicherheit des Sprechers das periphrastische Futur genutzt wird. Letztendlich kommt sie zu dem Schluss, dass der wichtigste Faktor bei der Verwendung der beiden Zukunftsformen die Unterscheidung zwischen geschriebenen und gesprochenen Texten ist.⁹⁷⁸ Der zweite Faktor ist die Sicherheit beziehungsweise Unsicherheit über das Eintreten des besprochenen Ereignisses. Nicht nur die hauptsächlichste Verwendung des periphrastischen Futurs zusammen mit der ersten Person, sondern auch in Kombination mit einer geringen zeitlichen Distanz können durch

⁹⁷⁴ Vgl. Sedano (2007): op. cit., S. 135.

⁹⁷⁵ Vgl. Sedano (2007): op. cit., S. 136.

⁹⁷⁶ Vgl. Sedano (2007): op. cit., S. 137-138.

⁹⁷⁷ Vgl. Sedano (2007): op. cit., S. 138.

⁹⁷⁸ Vgl. Sedano (2007): op. cit., S. 141.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

die Sicherheit erklärt werden, da näher an der Gegenwart liegende Ereignisse mit höherer Wahrscheinlichkeit vorhergesagt werden können als weiter entfernte und daher die Wahl auf das periphrastische Futur fällt.⁹⁷⁹

Diese Studie kann exemplarisch dazu herangezogen werden, um die Bedeutung von Korpusanalysen für den Fremdsprachenunterricht zu veranschaulichen. Anhand derartiger Untersuchungen werden klare Tendenzen im Sprachgebrauch ermittelt, welche für die Auswahl der Vokabel- und Grammatikeinheiten in Lehrbüchern berücksichtigt werden sollten. So legt die eben beschriebene Studie nahe, dass diese beiden Hauptfaktoren – die Mündlichkeit/Schriftlichkeit und die Sicherheit/Unsicherheit – auch in das jeweilige Grammatikkapitel der Lehrbücher aufgenommen werden sollten. Zudem ist die Unterscheidung im Hinblick auf die Sicherheit des Eintretens der Handlung derer zwischen dem *will-future* und dem *going-to-future* im Englischen sehr ähnlich. Auch dieser fächerübergreifende Kommentar ist für Schüler sehr hilfreich und erleichtert ihnen die Einordnung.

Im Folgenden wird dargestellt, wie beide Futurausdrücke in aktuellen Schulgrammatiken beschrieben werden, wobei weitere Angaben, welche sich nicht auf relevante Nutzungskontexte beziehen, wie beispielsweise die Verwendung des morphologischen Futurs für Ge- und Verbote oder der Gebrauch des periphrastischen Futurs zum Ausdruck einer Aufforderung in der 1. Person Plural, nicht angeführt werden. In der *¡Gramática! de la lengua española* von Klett wird angegeben, dass das *futuro simple*, also das morphologische Futur, genutzt wird, um die Zukunft auszudrücken und um Vermutungen zu äußern, die sich auf die Gegenwart beziehen, während man mit der Konstruktion *ir a + Infinitiv* über Absichten und Pläne in der Zukunft spricht. Hier wird zudem darauf hingewiesen, dass man in derartigen Situationen im Französischen das Verb *aller* verwendet, allerdings direkt mit dem Infinitiv.⁹⁸⁰ In der spanischen Grammatik für die Mittel- und Oberstufe von Cornelsen wird das *futuro simple* für mögliche Handlungen oder Zustände in der Zukunft, für Vermutungen oder Zweifel verwendet, während das *futuro inmediato* (*ir a + infinitivo*) beschreibt, was in der nahen Zukunft geschehen wird und ausdrückt, dass eine Handlung geplant oder ziemlich sicher ist.⁹⁸¹

⁹⁷⁹ Vgl. Sedano (2007): op. cit., S. 142-143.

⁹⁸⁰ Vgl. Dorn, Rudolf/ Navarro González, Javier/ Strathmann, Jochen (2012): *¡Gramática! de la lengua española*. Stuttgart: Klett, S. 75-76.

⁹⁸¹ Vgl. Rathsam / Schleyer (2014): op. cit., S. 205, 207.

Die *Lerngrammatik Spanisch* von Diesterweg gibt an, dass das *futuro simple* zum Ausdruck von zukünftigen Handlungen sowie von Vermutungen gebraucht wird. Die Konstruktion *ir a + Infinitiv* wird hier separat unter den Verbalperiphrasen geführt, ohne Verweis auf das *futuro simple*. Es wird angegeben, dass es sehr häufig gebraucht wird, vor allem in der Umgangssprache. Laut dieser Grammatik bedeutet *ir a + Infinitiv* in vielen Fällen nicht nur nahe Zukunft, sondern Zukunft schlechthin, wobei die Verbalperiphrase hauptsächlich im Präsens oder Imperfekt verwendet wird. Zudem drückt sie bisweilen auch eine Absicht aus, insbesondere wenn sie im Imperfekt steht.⁹⁸² In der praktischen Grammatik der spanischen Sprache des Gottfried Egert Verlags wird das Futur I, hier auch *futuro imperfecto* genannt, zur Bezeichnung in der Zukunft liegender Handlungen und Zustände sowie zum Ausdruck einer Vermutung in der Gegenwart verwendet. Hier wird jedoch zusätzlich darauf verwiesen, dass das Futur I in der Umgangssprache häufig durch die Verbalperiphrase *ir a + Infinitiv* ersetzt wird.⁹⁸³ Im entsprechenden Kapitel wird erläutert, dass diese Verbalperiphrase einen unmittelbar bevorstehenden Vorgang bezeichnet und häufig als Ersatz für das Futur I verwendet wird.⁹⁸⁴ Anhand dieses exemplarischen Einblicks in spanische Grammatiken ist ersichtlich, dass zwar die Unterscheidung Sicherheit/Unsicherheit über das Eintreten der Aussage meist ersichtlich wird, der Unterschied zwischen Mündlichkeit und Schriftlichkeit bis auf den Hinweis zum häufigen Gebrauch der Verbalperiphrase in der Umgangssprache in den Grammatiken von Diesterweg und des Gottfried Egert Verlags jedoch nicht erwähnt wird. Außerdem wird in der *¡Gramática! de la lengua española* von Klett zwar der Vergleich hergestellt zwischen *ir a + Infinitiv* und dem französischen *aller + Infinitiv*, doch es findet sich in keiner Grammatik ein Verweis auf die englische Unterscheidung der Zukunftszeiten, die der spanischen zumindest im Hinblick auf die vermutete Wahrscheinlichkeit des Eintretens der zukünftigen Handlung ähnelt. Im Anfangsunterricht, und damit in den Lehrwerken der ersten beiden Lernjahre, müssen diese Unterscheidungskriterien noch nicht in aller Ausführlichkeit dargestellt werden, in allgemeinen spanischen Grammatiken, welche auch von Oberstufenschülern genutzt werden, könnte darauf aber durchaus hingewiesen werden. Der Vergleich mit dem

⁹⁸² Vgl. Süß, Kurt/ Pérez, Petronilo/ Ruipérez, Germán (2003): *Lerngrammatik Spanisch*. Braunschweig: Diesterweg, S. 37-38, 108-109.

⁹⁸³ Vgl. Reumuth, Wolfgang/ Winkelmann, Otto (2011): *Praktische Grammatik der spanischen Sprache*. Wilhelmsfeld: Gottfried Egert Verlag, S. 218.

⁹⁸⁴ Vgl. Reumuth/ Winkelmann (2011): op. cit., S. 328-329.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

Englischen würde hingegen allen Schülern helfen.

3.3.2.2.3. Studie zur Auswahl des Vokabulars für Lehrwerke

In Bezug auf den Wortschatz in Lehrbüchern ist zu sagen, dass einige Einheiten des Lernwortschatzes in der außerschulischen Welt so selten sind, dass die Schüler ihnen im spanischen Sprachraum oder in spanischen Medien schlichtweg nie begegnen werden. Daher sollten Häufigkeitsanalysen bei der Gestaltung von Lehrwerken durchaus berücksichtigt werden. Beim Aufbau des Vokabelteils kann allerdings nicht nur auf die Häufigkeit geachtet werden, sondern es werden auch andere, nicht so häufig verwendete Begriffe in ein Lehrwerk aufgenommen, um ein ganzheitliches Bild der Sprache und der spanischen Lebenswelt zu schaffen. Der Großteil des Vokabulars sollte jedoch im spanischen Sprachraum so häufig Anwendung finden, dass die Schüler diesen Lexemen wirklich begegnen könnten und sie somit aufgrund der tatsächlichen Anwendbarkeit des Gelernten motivierter und konzentrierter lernen können. Außerdem ist die Zeitersparnis zu berücksichtigen. Sofern nur die häufigsten Lexeme des Spanischen zu erlernen sind, werden seltene Ausdrücke weggelassen, ohne dass das Verständnis spanischer Texte beeinträchtigt wird. Diese Berücksichtigung der Textabdeckung durch das gelernte Vokabular führt zu effizienterem und wirklichkeitsnahem Lernen.

Wie in Kapitel 3.3.2. bisher bereits zu sehen war, können Korpora grundsätzlich bei der Entscheidung helfen, was unterrichtet werden soll, also in Bezug auf den Lehrplan, sowie bei der Erstellung von geeigneten Materialien, um diesen Lehrplan zu realisieren. Die im Internet für jeden frei zugänglichen Textsammlungen stellen daher im Bereich der Wortschatzarbeit durchaus eine Bereicherung für den Fremd- und Zweitspracherwerb dar. Dabei spielt nicht nur die Kontextualisierung der Einheiten, sondern auch deren realer Gebrauch durch Muttersprachler eine Rolle. Die in gängigen Schul- und Textbüchern verwendeten Lexeme befinden sich im Umfeld künstlich geschaffener Texte, die nicht von Muttersprachlern geschrieben oder gesprochen, sondern von Didaktikern und muttersprachlichen Autoren speziell zu Unterrichtszwecken entwickelt wurden. Das in diesen Texten verwendete Vokabular wird also gemäß der Progression der Texte gelernt und mithilfe auf die Textinhalte abgestimmter Übungen verinnerlicht. Laut Davies/Face spiegelt diese Auswahl an Lexemen aber vermutlich nicht die tatsächlich am häufigsten verwendeten Einheiten der Sprache wider. Die Problematik dieses Sachverhalts liegt darin begründet, dass Lernende einer

Fremdsprache, nach Meinung der Autoren, diejenigen Lexeme der anderen Sprache erlernen sollten, die ihnen im Sprachkontakt am wahrscheinlichsten begegnen werden. In spanischsprachigen Textbüchern wird darauf jedoch bislang keine Rücksicht genommen, soweit die Autoren. Davies/Face sind der Meinung, das Ziel eines Textbuchs sollte es sein, die Schüler auf die 'reale Welt' vorzubereiten, so wie sie von Daten aus einem akkuraten Korpus mit Sprachmaterial aus der 'realen Welt' repräsentiert wird. Es sollte nicht das Ziel sein, nur die Bereiche und Sprachfunktionen der 'Textbuchsprache' zu beherrschen.⁹⁸⁵

Ein großes Problem im Bereich der Lehrmaterialien des Spanischen war bisher die mangelnde Verfügbarkeit von Korpora und frequenzbasierten Wörterbüchern. Dank der seit geraumer Zeit verfügbaren repräsentativen Korpora im Bereich des Englischen genießen Lerner dieser Sprache schon seit längerem ein großes Angebot an exzellenten Lehrmitteln, die auf realistischen Daten aus repräsentativen Korpora basieren. Beispiele hierfür sind, wie in Kapitel 3.3.2.1. erwähnt, vollständig korpusbasierte Wörterbücher sowie Grammatiken. Außerdem gibt es wortschatzorientierte Lehrbücher für Fremdsprachenlerner, welche vollständig auf Korpusdaten basieren (beispielsweise Schmitt and Schmitt 2005). Für viele andere Sprachen, wie eben auch das Spanische, gibt es leider kein derart ausgebautes Angebot an korpusbasierten Materialien.⁹⁸⁶

In ihrer Studie untersuchten Davies/Face sechs Spanischlehrwerke für englischsprachige Schüler, von denen keines mit Vokabular arbeitet, welches auf korpusbasierten Häufigkeitsanalysen beruht. Sie gaben außerdem an, dass dies auch bei einigen anderen Textbüchern, welche sie nicht derart genau untersuchten, der Fall wäre. Ihrer Meinung nach liegt diese geringe Berücksichtigung der wirklichen Häufigkeit der Lexeme an zwei Faktoren. Zum einen geben sie die vorherrschende Kultur der Lehrwerkverlage an. Zum anderen, und diese Variante ist laut Davies/Face die wahrscheinlichere, rechtfertigen sie das Verhalten der Textbuchersteller durch die geringe Anzahl an großen, repräsentativen Korpora. Folglich gäbe es auch wenige, wenn überhaupt wirklich repräsentative Häufigkeitswörterbücher, auf denen die Textbuchersteller ihre Materialien aufbauen könnten.⁹⁸⁷

⁹⁸⁵ Vgl. Davies/ Face (2006): op. cit., S. 132-133.

⁹⁸⁶ Vgl. Davies/ Face (2006): op. cit., S. 133.

⁹⁸⁷ Vgl. Davies/ Face (2006): op. cit., S. 133.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

Für ihren Vergleich zwischen dem Vokabular im CdE:Orig (siehe Kapitel 2.1.1.3.) und dem Vokabular in Spanischlehrbüchern analysierten Davies/Face das zu erlernende Vokabular von sechs im Spanischunterricht häufig verwendeten Lehrbüchern. Dabei zogen sie je drei Lehrbücher aus dem ersten und drei aus dem zweiten Lernjahr heran. Ziel der Studie von Davies/Face war es, herausfinden, inwieweit das vom Lehrwerk abgedeckte Vokabular mit den häufigsten Lexemen des CdE:Orig und somit des *Routledge Frequency Dictionary of Spanish* (siehe Kapitel 3.3.2.1.) übereinstimmt. In ihrer Analyse verglichen sie die Abdeckung der frequentesten Lexeme nach der Wortart, dem Niveau des Textbuchs – also dem ersten oder zweiten Lernjahr – und den einzelnen Lehrwerken.⁹⁸⁸

Zunächst führten Davies/Face quantitative Untersuchungen durch, indem sie den Prozentsatz der Lexeme in den Textbüchern verglichen, die auch im Frequenzwörterbuch vorkommen. Dabei kamen sie zu dem Ergebnis, dass ungefähr 93% der 500 häufigsten Einheiten des Frequenzwörterbuchs in mindestens einem der sechs Textbücher erscheint. Dieser Prozentsatz verringert sich allerdings auf 55% bei den Lexemen auf den Rängen 2001 bis 2500 und sogar auf 29% bei den Lexemen auf den Rängen 4501 bis 5000. In Bezug auf die Wortart stellten sie fest, dass die beste Abdeckung im Bereich der Substantive liegt und die Adverbien am wenigsten gut abgedeckt sind.⁹⁸⁹

Im Vergleich der Lehrbücher untereinander fanden sie heraus, dass die Lehrwerke des ersten Lernjahrs, wie erwartet, einen höheren Prozentsatz der häufigsten Lexeme abdecken als die Werke des zweiten Lernjahrs. Im Durchschnitt behandeln 1,98 von 3 Lehrwerken des ersten Jahrs die 500 häufigsten Einheiten des Frequenzwörterbuchs im Vergleich zu 0,95 von 3 Lehrwerken des zweiten Lernjahrs. Dieses Ergebnis überrascht allerdings wenig, da für gewöhnlich im ersten Lernjahr das Basisvokabular vermittelt wird, auf welches in folgenden Lernjahren zurückgegriffen wird. Dieses Vokabular muss somit nicht erneut aufgeführt werden. Paradoxerweise werden die weniger häufigen Begriffe, also diejenigen, die im Wörterbuch auf Rang 3500 bis 5000 stehen, nicht verstärkt im zweiten Lernjahr behandelt. Die Autoren erklärten sich diesen Sachverhalt dadurch, dass sich die Lehrwerke des zweiten Lernjahrs verstärkt mit aktuellen Problemen wie dem Umweltschutz oder dem Kampf gegen ungerechte soziale Zustände

⁹⁸⁸ Vgl. Davies/ Face (2006): op. cit., S. 135-137.

⁹⁸⁹ Vgl. Davies/ Face (2006): op. cit., S. 137-138.

beschäftigen. Die Bearbeitung dieser Themen sei zwar vorbildlich, das in den betreffenden Lektionen gelernte Vokabular sei allerdings, zumindest außerhalb dieser begrenzten Bereiche, von geringem Wert für die mündliche und schriftliche Sprachproduktion. Zudem fiel auf, dass die einzelnen Lehrwerke sich stark in der Summe der zu erlernenden Einheiten unterscheiden. Im höchsten Fall sind es 3217 und im geringsten nur 523 Lexeme. Die Autoren untersuchten außerdem das Verhältnis zwischen Lehrwerkvokabular und Frequenzwörterbuch, indem sie die Gesamtanzahl der Lexeme des Lehrwerks mit der Anzahl der darin enthaltenen Einheiten der 5000 häufigsten Lexeme des Frequenzwörterbuchs verglichen. Bei diesem Vergleich lag die höchste Korrespondenz bei 50%, die niedrigste lediglich bei 10%. In diesem Fall sind 90% der zu erlernenden Einheiten nicht in der Liste der 5000 häufigsten Lexeme enthalten. Dieses Ergebnis führten Davies/Face wiederum auf die Annahme der Lehrwerkautoren zurück, die Schüler wären im zweiten Lernjahr bereits soweit fortgeschritten, dass sie auf die Einführung des Basisvokabulars verzichten und anstatt dessen aktuelle soziale Probleme behandeln könnten.⁹⁹⁰

Nach diesen quantitativen Untersuchungen versuchten die Autoren, das Vokabular der Lehrwerke qualitativ auszuwerten. Dabei untersuchten sie zunächst diejenigen Begriffe, die im Vergleich zum Frequenzwörterbuch in den Lehrwerken unterrepräsentiert sind. Hierfür suchten sie diejenigen Einheiten, die im Frequenzwörterbuch zu den 1000 häufigsten Lexemen zählen, im Lehrbuch aber nicht aufgelistet werden. Aufgeteilt auf die Wortarten kommen sie dabei zu folgendem Ergebnis:

Wortart	Substantiv	Verb	Adjektiv	Adverb
Anzahl	51	34	28	14
Verhältnis zur Anzahl der Einheiten dieser Wortart unter den häufigsten 1000 Lexemen	12,8% aller Substantive	12,1%	16,9%	19,4%

Abbildung 78: Anzahl der in Lehrwerken unterrepräsentierten Lexeme je Wortart.⁹⁹¹

Während die größte Gruppe der unterrepräsentierten Einheiten die Substantive bilden, sind diese prozentual gesehen nicht die am meisten unterrepräsentierte Wortart, sondern die Adverbien. Weit wichtiger ist allerdings die Frage, welche Substantive,

⁹⁹⁰ Vgl. Davies/ Face (2006): op. cit., S. 138-140.

⁹⁹¹ Vgl. Davies/ Face (2006): op. cit., S. 140.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

Verben, Adjektive und Adverbien unterrepräsentiert sind. Eine genauere Untersuchung der Lexeme ergab, dass es sich dabei um relativ abstrakte Begriffe handelt, die Kognate des Englischen sind, wie beispielsweise *momento*, *situación* oder *especial*. Allerdings kann die Begründung des generellen Verzichts auf Kognate widerlegt werden, da andere Kognate wie *animal*, *color* oder *social* sehr wohl in den Vokabellisten enthalten sind. Deshalb kamen Davies/Face zu dem Schluss, dass in den Lehrwerken des ersten und zweiten Lernjahrs vor allem abstrakte Begriffe unterrepräsentiert sind.⁹⁹²

Ebenso untersuchten die Autoren überrepräsentierte Einheiten. Dabei legten sie fest, dass all diejenigen Lexeme überrepräsentiert sind, die in mindestens zwei der sechs Textbücher vorkommen, im Korpus aber nur zwischen zwei und 100-mal auftreten. Bei der Überrepräsentation gaben Davies/Face die genau gegenteilige Begründung an: Die hier erscheinenden Begriffe sind fast ausschließlich konkreter Natur. So finden sich unter ihnen beispielsweise die Lexeme *butaca* (2 Mal im Korpus), *patinar* (28) und *chistoso* (46). Es gibt nur einige wenige überrepräsentierte Lexeme, die eher abstrakt sind. Zu ihnen zählen Lexeme wie *promedio* (5), *idealista* (70) und *desperdiciar* (84).⁹⁹³

Um nun die scheinbare Verbindung zwischen unterrepräsentierten Lexemen und einem hohen Abstraktionsgrad sowie überrepräsentierten Lexemen und einem niedrigen Abstraktionsgrad zu erklären, führten Davies/Face die Rolle prädefinierter semantischer Felder an. Ihrer Meinung nach sei es offensichtlich, dass Lehrwerkautoren semantische Felder nutzten, um Vokabular, Textpassagen und andere Übungen zu gruppieren. Dies sei offensichtlich bei der Betrachtung einzelner Kapitel des Textbuchs, welche sich mit Themen wie Universität, Familie, Einkaufen, Zuhause, Wetter, Umwelt, Essen, Gesundheit und Freizeit beschäftigen. Konkrete Konzepte fügen sich dabei besser in die semantischen Felder, weshalb sie überrepräsentiert werden, wenn Textbuchautoren mehr Vokabular für einzelne Kapitel benötigen.⁹⁹⁴

Davies/Face sprechen sich zwar nicht für einen Verzicht auf semantische Felder in Lehrwerken aus, sie propagieren aber verstärkt die Einbeziehung ihrer Forschungsergebnisse in die Entwicklung von Spanischlehrwerken. Würden Schüler nur die 1000 häufigsten Lexeme des Spanischen lernen, so verstünden sie bereits 76-80% des geschriebenen Spanisch, je nachdem, ob es sich um fiktive oder nicht-fiktive Texte

⁹⁹² Vgl. Davies/ Face (2006): op. cit., S. 140-141.

⁹⁹³ Vgl. Davies/ Face (2006): op. cit., S. 140-141.

⁹⁹⁴ Vgl. Davies/ Face (2006): op. cit., S. 141-142.

handle, und sogar 88% des gesprochenen Spanisch. Die untersuchten Textbücher enthalten jedoch nur 10-50% der häufigsten spanischen Lexeme. Die Autoren hoffen, dass demnächst auch für das Spanische korpusbasierte Lehrwerke erstellt werden, so wie es im Bereich des Englischen als Fremdsprache bereits der Fall ist.⁹⁹⁵

3.3.2.3. Erstellung von Sprachtests

Der Bereich der Sprachtests profitiert ebenso vom korpusbasierten Ansatz. Korpora werden für die Testentwicklung, -zusammenstellung und -auswahl, die Testpräsentation, die Erfassung der Antworten sowie die Berechnung und Übermittlung der Ergebnisse genutzt. Alderson sieht große Vorteile darin, Tests auf der Basis von realen Sprachdaten zu erstellen sowie Einschätzungen zu Fähigkeiten, Wissen und Leistung von Schülern anhand dieser Daten zu treffen.⁹⁹⁶

Zur Zeit der Veröffentlichung seines Aufsatzes im Jahr 1996 war allerdings noch nicht klar, ob diese Vorteile in die Praxis umgesetzt werden könnten. Heute kann man jedoch anhand der Gleichwertigkeit des computerbasierten TOEFL-Tests und dessen Pendant auf Papier erkennen, dass diese Zweifel zufriedenstellend ausgeräumt wurden. Korpora sind in den vergangenen Jahren von Testservices beispielsweise als Archiv für Testskripte, zur Entwicklung von Testmaterialien, zur Optimierung von Testverfahren, zur Verbesserung der Qualität der Testbewertung sowie zur Validierung und Standardisierung von Tests verwendet worden.⁹⁹⁷

Im englischsprachigen Bereich wurden Korpora bereits mehrfach für Sprachtests genutzt. So verwendete Coniam⁹⁹⁸ Daten zur Worthäufigkeit für die automatische Erstellung von Lückentests und Kaszubski/Wojnowska⁹⁹⁹ erstellten ein korpusbasiertes Computerprogramm für satzbasierte Übungen, den TestBuilder. Taylor/Barker (2008) bieten einen Überblick über die Entwicklung und Nutzung von Korpora auf dem Gebiet

⁹⁹⁵ Vgl. Davies/ Face (2006): op. cit., S. 142-143.

⁹⁹⁶ Vgl. Alderson, Charles (1996): „Do corpora have a role in language assessment?“ In: Thomas, J./ Short, M. (Hrsg.) (1996): *Using corpora for language research*. London: Longman, S. 258–259.

⁹⁹⁷ Vgl. McEnery/ Xiao (2010): op. cit., S. 369.

⁹⁹⁸ Vgl. Coniam, David (1997): „A preliminary inquiry into using corpus word frequency data in the automatic generation of English language cloze tests“. In: *CALICO Journal* 16 (2–4), S. 15–33.

⁹⁹⁹ Vgl. Kaszubski, Przemysław/ Wojnowska, Aleksandra (2003): „Corpus-informed exercises for learners of English: The TestBuilder program“. In: Oleksy, E./ Lewandowska-Tomaszczyk, B. (Hrsg.) (2003): *Research and scholarship in integration processes: Poland–USA–EU*. Łódź: Łódź University Press, S. 337–354.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

der Sprachtests.¹⁰⁰⁰ Von der aktuellen Brisanz dieser Thematik in der angewandten Linguistik und im Fremdspracherwerb auch im Hinblick auf die Verwendung von Lernerkorpora zeugen Barker/Salamoura/Saville (2015)¹⁰⁰¹ sowie Callies/Götz (2015)¹⁰⁰². In Callies/Götz (2015) werden die Vorzüge und Möglichkeiten der Nutzung von Lernerkorpora für die Prüfung und Bewertung von schriftlichen und mündlichen Fremdsprachenkenntnissen betont. Es werden neue Korpusressourcen für die Analyse der Sprachkenntnisse, Programme für die automatische Sprachbewertung sowie Methoden der Selbstbewertung anhand von Lernerkorpora vorgestellt. In erster Linie handelt es sich hierbei jedoch um Englisch als Fremdsprache. Higgins/Ramineni/Zechner (2015) beschäftigen sich mit der Nutzung von Lernerkorpora für die automatisierte Auswertung, insbesondere im Hinblick auf offene Aufgabenstellungen, wie das Verfassen eines Essays.¹⁰⁰³

Alonso-Ramos kritisiert im Hinblick auf Sprachtests, dass obwohl das *Instituto Cervantes* und die *Universidad de Salamanca* Mitglieder der *Association of Language Testers in Europe* (ALTE) sind, die Präsenz des Spanischen in den Konferenzen rar ist. Lernerkorpora sind jedoch von Bedeutung für die Entwicklung von Sprachtests, sowohl im Hinblick auf die Einstufung als auch auf die Bewertung der Sprachkenntnisse. Da immer mehr Studenten Zertifikate über ihre Sprachkenntnisse benötigen, um beispielsweise im Ausland zu studieren, ist es laut Alonso-Ramos nicht weit hergeholt, über den Nutzen automatisierter Bewertungsmethoden nachzudenken. Sie bedauert die geringe Anzahl an Forschungsvorhaben, welche sich mit spanischen Sprachtests beschäftigen, und verweist auf *Avalingua* von Gamallo et al. (2015). Alonso-Ramos schlägt vor, dieses Programm auf das Spanische zu übertragen.¹⁰⁰⁴ Hierbei handelt es sich um ein Programm zur automatischen Fehleranalyse für die galicische Sprache. Mittels

¹⁰⁰⁰ Vgl. Taylor, Lynda/ Barker, Fiona (2008): „Using Corpora for Language Assessment“. In: Shohamy, E./ Hornberger N.H. (Hrsg.) (2008): *Encyclopedia of Language and Education. Language Testing and Assessment*. Second Edition, Vol. 7, New York: Springer, S. 241-254.

¹⁰⁰¹ Vgl. Barker, Fiona/ Salamoura, Angeliki/ Saville, Nick (2015): „Learner Corpora and Language Testing“. In: Granger, S./ Gilquin, G./ Meunier, F. (Hrsg.) (2015): *The Cambridge Handbook of Learner Corpus Research*. Cambridge: Cambridge University Press, S. 511-534.

¹⁰⁰² Vgl. Callies, Marcus/ Götz, Sandra (Hrsg.) (2015): *Learner Corpora in Language Testing and Assessment*. Amsterdam: John Benjamins.

¹⁰⁰³ Vgl. Higgins, Derrick/ Ramineni, Chaitanya/ Zechner, Klaus (2015): „Learner Corpora and Automated Scoring“. In: Granger, S./ Gilquin, G./ Meunier, F. (Hrsg.) (2015): *The Cambridge Handbook of Learner Corpus Research*. Cambridge: Cambridge University Press, S. 587-604.

¹⁰⁰⁴ Vgl. Alonso-Ramos (2016): op. cit., S. 18.

Avalingua werden grammatikalische Fehler in geschriebener Sprache erkannt und kategorisiert, Textproduktionen bewertet sowie Vorschläge zur Verbesserung der sprachlichen Fähigkeiten gemacht. Das Programm wurde mithilfe zweier Lernerkorpora überprüft und erreichte eine Trefferquote von 91% für die Erkennung verschiedener Arten grammatikalischer Fehler, inklusive Rechtschreib-, lexikalischer und syntaktischer Fehler.¹⁰⁰⁵

Wie zuvor erwähnt, werden Korpora bereits umfangreich auf indirekte Weise verwendet, selbst wenn sich viele Nutzer dessen nicht bewusst sind, beispielsweise bei der Verwendung eines korpusbasierten Wörterbuchs. Es gibt einige Nachschlagewerke, Lehrpläne, Unterrichtsmaterialien und Sprachtests, die auf Korpusdaten basieren und deren Nützlichkeit niemand mehr abstreiten würde. Da die Erstellung korpusbasierter Referenzwerke, Lehrbücher und zentraler Sprachtests, geschweige denn des Lehrplans, jedoch nicht in den Arbeitsbereich des einzelnen Lehrers fällt, sollen die bisherigen Ausführungen zu diesem Bereich genügen, da vor allem die direkte Nutzung der Korpora für die Lehrkräfte von Belang ist.

3.3.3. Unterrichtsorientierte Korpora

Unterrichtsorientierte Korpora sind besonders nützlich für den Unterricht von LSP (*Languages for Specific Purposes*) sowie für Untersuchungen im Erst- und Zweitspracherwerb. Diese Korpora können in der Sprachpädagogik sowohl direkt als auch indirekt verwendet werden.¹⁰⁰⁶ Alonso-Ramos gibt an, dass die Forschung zu Lernerkorpora, auch LCR (*Learner Corpus Research*) genannt, aus einer Kombination der drei Disziplinen Korpuslinguistik, Zweitspracherwerb und Fremdsprachendidaktik entstanden ist.¹⁰⁰⁷

Unter *Languages for Specific Purposes* verstehen McEnery/Wilson „the domain specific teaching of a language for a particular area of application, for example, the teaching of medical English to medical students“. Diese Art von Korpora können dazu

¹⁰⁰⁵ Vgl. Gamallo Otero, Pablo/ Garcia, Marcos/ del Río, Iria/ González López, Isaac (2015): „Avalingua. Natural language processing for automatic error detection“. In: Callies, M./ Götz, S. (Hrsg.) (2015): *Learner Corpora in Language Testing and Assessment*. Amsterdam: John Benjamins, S. 35-58.

¹⁰⁰⁶ Vgl. McEnery/ Xiao (2010): op. cit., S. 371.

¹⁰⁰⁷ Vgl. Alonso-Ramos (2016): op. cit., S. 16.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

genutzt werden, um verschiedene Arten bereichsspezifischer Materialien bereitzustellen (unter anderem quantitative Auskünfte zum Vokabular und dessen Gebrauch), welche auf Lerner dieser speziellen Sprachbereiche besser zugeschnitten sind als Informationen aus allgemeinen Korpora.¹⁰⁰⁸

Tolchinsky führt aus, dass bis vor kurzem die meisten Korpora Muster der normativen Sprache beinhalten, derzeit jedoch mehr und mehr Korpora mit Texten aus dem Internet zusammengestellt werden, welche in offenen und spontanen Situationen entstanden und in denen die normativen Konventionen der geschriebenen Sprache verändert sind. Diese Veränderung findet sich auch in Korpora mit Texten von Fremdsprachenlernern, von Schülern, die gerade das Schreiben erlernen, oder von jeglichen Sprechern in einer informellen Kommunikationssituation. Prinzipiell könnte jeder Sprachlehrer ein Lernerkorpus zusammenstellen und es didaktisch nutzen. Bisher wurden Lernerkorpora hauptsächlich im Zweitspracherwerb genutzt, sie können allerdings auch für den Erstspracherwerb oder in anderen Bereichen verwendet werden.¹⁰⁰⁹ Osborne spricht hierbei von zwei verschiedenen Ansätzen in Bezug auf die Verwendung von muttersprachlichen im Gegensatz zu Lernerkorpora. Beim sogenannten *top-down* Ansatz, welcher meist beim *Data-Driven Learning* (siehe Kapitel 3.3.1.2.) angewandt wird, werden Daten von Muttersprachlern als Belege dafür verwendet, wie die Zielsprache strukturiert ist, sowie als Vorbild für die Sprachproduktion der Schüler. Als Vorteil dieser Vorgehensweise nennt er die Authentizität der Daten sowie die Veranschaulichung von Mustern und Kollokationen (zum Begriff der Kollokation siehe Kapitel 4.1.1.). Bei der Arbeit mit Lernerkorpora findet hingegen der *bottom-up* Ansatz Anwendung, da hier von den Lernerdaten ausgegangen wird. Dies bietet den Vorteil, dass zunächst analysiert wird, was die Lerner mit der Sprache ausdrücken möchten, um ihnen daraufhin falsche Gebrauchsmuster bewusst zu machen und ihnen zu helfen, diese zu verbessern.¹⁰¹⁰

¹⁰⁰⁸ McEnery/ Wilson (1996): op. cit., S. 105.

¹⁰⁰⁹ Vgl. Tolchinsky (2014): op. cit., S. 15.

¹⁰¹⁰ Vgl. Osborne, John (2004): „Top-down and Bottom-up Approaches to Corpora in Language Teaching“. In: Connor, U./ Upton, T. A. (Hrsg.) (2004): *Applied Corpus Linguistics. A Multidimensional Perspective*. Amsterdam [u.a.]: Rodopi, S. 252-253.

Die Forschung zu Lernerkorpora ist noch ein recht junges Forschungsfeld. Sie wird erst seit den späten 1980er Jahren betrieben.¹⁰¹¹ Laut Alonso-Ramos steckt sie besonders im Hinblick auf spanische Lernerkorpora sozusagen noch in den Kinderschuhen.¹⁰¹² Fernández Pérez und ihre Kollegen untersuchten, wie in der Einleitung zum dritten Kapitel angesprochen, die Aneignung der spanischen Sprache im Hinblick auf den Erstspracherwerb anhand des Korpus *Koiné*, eines Korpus der Kindersprache, und fanden dabei einige Merkmale, Elemente und Eigenschaften heraus, welche unterliegende Prinzipien der Sprachaneignung sowie der Kommunikationstaktik darstellen.¹⁰¹³

López-Ferrero verglich den Gebrauch parataktischer Konjunktionen in Textproduktionen von spanischen Schülern nach der Sekundarstufe, welche vor dem Beginn ihres Universitätsstudiums standen, mit demjenigen in Texten, welche von erfahrenen Spezialisten geschrieben wurden. Das spezifische Ziel ihrer Studie war die Beschreibung der syntaktischen, semantischen, pragmatischen und Diskurseigenschaften von parataktischen Konjunktionen wie *y/e*, *ni*, *o/u*, *bien ... bien* oder *pero*, wobei sie das in Kapitel 2.5.1.1. bereits beschriebene *Bwananet* des *Institut Universitari de Lingüística Aplicada* der *Universitat Pompeu Fabra* nutzte. Sie fand in ihrer Studie heraus, dass die Schüler Probleme hatten bei der Beherrschung der parataktischen Beziehung in Einheiten der Mikrostruktur (Satzebene) und der Makrostruktur (Textebene), bei der Nutzung der Konjunktionen zur Organisation von formalen, akademischen Diskursen sowie bei der Kombination der Konjunktionen mit anderen textuellen Einheiten. Zudem ist der semantische Wert dieser Konjunktionen im geplanten, geschriebenen Register ein anderer als im umgangssprachlichen, informellen Register. Da die Schüler keinen Unterschied zwischen diesen Registern machen und es ihnen an Planung fehlt, nutzen sie parataktische Konjunktionen zwangsläufig auf unangemessene Weise, um Informationen zu strukturieren.¹⁰¹⁴

Im Englischen wird dem bereichsspezifischen Sprachgebrauch und der beruflichen Kommunikation viel Aufmerksamkeit geschenkt, beispielsweise im Hinblick auf

¹⁰¹¹ Vgl. Granger, Sylviane (2002): „A bird’s-eye view of learner corpus research“. In: Granger, S./ Hung, J./ Petch-Tyson, S. (Hrsg.) (2002): *Computer learner corpora, second language acquisition and foreign language teaching*. Amsterdam: John Benjamins, S. 4.

¹⁰¹² Vgl. Alonso-Ramos (2016): op. cit., S. 3.

¹⁰¹³ Vgl. Fernández Pérez (2011): op. cit.

¹⁰¹⁴ Vgl. López-Ferrero (2007): op. cit., S. 173-194.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

Englisch für akademische Zwecke.¹⁰¹⁵ Thurstun/Candlin untersuchten den Gebrauch von Konkordanzen im Textproduktions- und Vokabelunterricht im Bereich des akademischen Englisch.¹⁰¹⁶ Das *English Language Institute* (ELI) der *University of Michigan* erstellte das *Michigan Corpus of Academic Spoken English* (MICASE), um nicht-muttersprachlichen, internationalen Studenten bei der Entwicklung und Verbesserung der kommunikativen Fähigkeiten, die sie zur Erlangung ihres Abschlusses benötigen, zu unterstützen.¹⁰¹⁷ Hinkel stellte Vergleiche zum Gebrauch von Zeiten, Aspekt und Passiv in akademischen Texten von Muttersprachlern und Fremdsprachenlernern an.¹⁰¹⁸ Nelson führte eine korpusbasierte Studie zum Vokabular von *Business English* und dazugehörigen Unterrichtsmaterialien durch.¹⁰¹⁹ Zudem sind parallele Konkordanzen nicht nur sinnvoll für den Übersetzungsunterricht, sie können auch beim sogenannten ‚reziproken Lernen‘ von Vorteil sein, wo zwei Sprachlerner mit unterschiedlichem muttersprachlichem Hintergrund zusammengebracht werden, um sich gegenseitig bei der Aneignung der Fremdsprache zu unterstützen.¹⁰²⁰ Derartige Studien zeigen, dass LSP-Korpora besonders nützlich sind für den Unterricht in LSP und beruflicher Kommunikation.¹⁰²¹

Im spanischsprachigen Bereich ist die Studie von Parodi/Gramajo zu nennen. Sie nutzten ihre Charakterisierung eines akademisch-beruflichen Diskursgenres zur Verfeinerung einer Texttypologie mit didaktischer Zielsetzung. Dabei bestimmten sie anhand von drei Korpora Texttypen, welche die Texte von drei Fachrichtungen einer *Secondary Middle School* aus funktional-kommunikativer und textueller Sicht

¹⁰¹⁵ Vgl. McEnery/ Xiao (2010): op. cit., S. 371.

¹⁰¹⁶ Vgl. Thurstun, Jennifer/ Candlin, Christopher N. (1997): *Exploring academic English: A workbook for student essay writing*. Sydney: NCELTR. Und: Thurstun, Jennifer/ Candlin, Christopher N. (1998): „Concordancing and the teaching of the vocabulary of academic English“. In: *English for Specific Purposes*, 17, S. 267–280.

¹⁰¹⁷ Vgl. Simpson, Rita/ Lucka, Bret/ Ovens, Janine (2000): „Methodological challenges of planning a spoken corpus with pedagogical outcomes“. In: Burnard, L./ McEnery, T. (Hrsg.) (2000): *Rethinking language pedagogy from a corpus perspective: papers from the Third International Conference on Teaching and Language Corpora*. Frankfurt am Main: Peter Lang, S. 43–49.

¹⁰¹⁸ Vgl. Hinkel, Eli (2004): „Tense, aspect and the passive voice in L1 and L2 academic texts“. In: *Language Teaching Research* 8(1), S. 5–29.

¹⁰¹⁹ Vgl. Nelson, Michael (2000): *A corpus-based study of business English and business English teaching materials*. PhD thesis. University of Manchester. Online Publikation: <http://users.utu.fi/micnel/thesis.html> (19.07.2017).

¹⁰²⁰ Vgl. Johns, Tim (1997): „Contexts: The background, development and trialling of a concordance-based CALL program“. In: Wichmann, A./ Fligelstone, S./ McEnery, T./ Knowles, G. (Hrsg.) (1997): *Teaching and language corpora*. Harlow: Longman, S. 100–115.

¹⁰²¹ Vgl. McEnery/ Xiao (2010): op. cit., S. 371.

charakterisieren. Sie konnten dabei innerhalb der beiden führenden Textprototypen (dem didaktischen und dem hochspezialisierten Prototyp) zwölf Texttypen identifizieren. Parodi/Gramajo fanden heraus, dass die jeweilige Fachrichtung (der maritime, der kommerzielle und der industrielle Sektor) Einfluss auf die Variabilität der Texttypen sowie auf die Anzahl der Texte, die in einem Sektor zirkulieren, hat. Generell ist die Verwendung didaktischer und hochspezialisierter Texttypen jedoch ausgeglichen, was für die Autoren sehr zufriedenstellend ist, da dies darauf hinweist, dass die Studenten schrittweise in das jeweilige Spezialgebiet eingeführt werden.¹⁰²²

Die Entwicklung und der Gebrauch von Lernerkorpora in der Sprachpädagogik und in der *Interlanguage*-Forschung stellen laut McEnery/Xiao eine der interessantesten Entwicklungen in der aktuellen korpusbasierten Sprachforschung dar.¹⁰²³ Der Begriff *Interlanguage* geht auf Selinker (1972) zurück, der feststellte, dass Lerner in einer bestimmten Situation das, was sie ausdrücken möchten, anders versprachlichen als Muttersprachler. Dieser Vergleich lässt auf die Existenz einer latenten psychologischen Struktur schließen, welche beobachtet werden kann, wenn man die Äußerungen eines Lerners untersucht, der versucht seine Gedanken in der Fremdsprache wiederzugeben. Sie zeigt sich nicht, wenn der Lerner formorientierte Aufgaben ausführt, wie bei mündlichen Drills im Klassenzimmer.¹⁰²⁴

Daten aus Lernerkorpora sind besonders nützlich für die Sprachpädagogik und die Zweitspracherwerbsforschung. Der Spracherwerb findet in den Köpfen der Lerner statt, kann daher nicht direkt beobachtet und muss aus einer psychologischen Perspektive untersucht werden. Wenn die Leistung der Lerner jedoch durch solch einen mentalen Prozess geformt und eingeschränkt wird, so stellt sie doch immerhin einen indirekten, beobachtbaren und empirischen Beweis für den Spracherwerbsprozess dar. Diese Nutzung des Produkts als Beweis für den Prozess muss laut McEnery/Xiao jedoch nicht zwingend weniger zuverlässig sein. Manchmal sei es der einzig praktikable Weg, um den Prozess zu untersuchen.¹⁰²⁵

In der Sprachpädagogik wurden die Implikationen der Lernerkorpora zum Zweck

¹⁰²² Vgl. Parodi/ Gramajo (2007): op. cit., S. 145-172.

¹⁰²³ Vgl. McEnery/ Xiao (2010): op. cit., S. 372.

¹⁰²⁴ Vgl. Selinker, Larry (1972): „Interlanguage“. In: *International Review of Applied Linguistics in Language Teaching*, 10 (3), S. 209-231.

¹⁰²⁵ Vgl. McEnery/ Xiao (2010): op. cit., S. 372.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

des Lehrplandesigns, der Entwicklung von Unterrichtsmaterialien und der Unterrichtsmethodik untersucht.¹⁰²⁶ Meunier beispielsweise gibt an, dass aus muttersprachlichen Korpora erhaltene Informationen zur Frequenz allein nicht ausreichend sind für die Erstellung von Lehrplänen und Unterrichtsmaterialien. Vielmehr muss eine Balance zwischen Häufigkeit, Schwierigkeit und pädagogischer Relevanz gefunden werden und genau hier kommt, ihrer Meinung nach, die Erforschung der Lernerkorpora ins Spiel, um die Bedeutung jedes dieser Kriterien gegeneinander abzuwägen.¹⁰²⁷

Die bisherigen Untersuchungen zur *Interlanguage*, welche auf Lernerkorpora basieren, beziehen sich auf die sogenannte *Contrastive Interlanguage Analysis* (CIA) von Granger, welche die Lernerdaten mit Daten von Muttersprachlern vergleicht. Die erste Art des Vergleichs zielt typischerweise darauf ab, den Grad des *Over-* und *Underuse* bestimmter linguistischer Merkmale in der Lernersprache zu ermitteln, während die zweite Art die Interferenz oder den Transfer aus der Muttersprache aufdecken soll. Es können auch Korpusdaten von Lernern mit unterschiedlichen Muttersprachen miteinander verglichen werden, um gemeinsame Eigenschaften des Spracherwerbs herauszufinden, die eben nicht unter dem Einfluss der jeweiligen Muttersprache stehen. Außerdem werden Lernerkorpora dazu verwendet, um die Reihenfolge der Aneignung bestimmter Morpheme zu untersuchen.¹⁰²⁸ Siepmann ermittelte anhand von Lernerkorpora Regelmäßigkeiten in der Interimsprache deutscher Lerner des Englischen. Zum einen ließen sich mit Lernerkorpora seiner Meinung nach rekurrente Valenzfehler nachweisen, zum anderen könnten damit auch Erkenntnisse zur Über- und Unterrepräsentation bestimmter Phänomene ermittelt werden.¹⁰²⁹

Ein Lernerkorpus kann einerseits Aufschluss über frequente Fehler von Schülern geben und andererseits eine nützliche Grundlage zur Erstellung von Übungen bilden. Der Lehrer kann ein Korpus mit von Schülern geschriebenen Texten erstellen, sofern diese Texte bereits digital vorliegen, und er kann anhand dieses Korpus für sein weiteres

¹⁰²⁶ Vgl. Keck, Casey Mari (2004): „Corpus linguistics and language teaching research: Bridging the gap“. In: *Language Teaching Research*, 8 (1), S. 99.

¹⁰²⁷ Vgl. Meunier, Fanny (2002): „The pedagogical value of native and learner corpora in EFL grammar teaching“. In: Granger, S./ Hung, J./ Petch-Tyson, S. (Hrsg.) (2002): *Computer learner corpora, second language acquisition and foreign language teaching*. Philadelphia: John Benjamins, S. 123.

¹⁰²⁸ Vgl. Granger (2002): op. cit., S. 3–33.

¹⁰²⁹ Vgl. Siepmann (2009): op. cit., S. 325–326.

Vorgehen ermitteln, welche Unterrichtseinheiten noch nicht verstanden worden sind und daher noch einmal vertieft werden sollten. Die mit Konkordanzen aus diesem Korpus erstellten Übungen können den Schülern dazu dienen, nach der Vertiefung problematischer Grammatikbereiche oder Vokabelfelder deren korrekten Gebrauch einzuüben und Fehler der anderen Schüler zu vermeiden. James/Garrett schreiben hierzu “once we realize what we *do* know, we are able to identify what it is that we *need* to know”.¹⁰³⁰

Lernerkorpora können zudem der Anfertigung von Tests dienen. In diesem Fall werden in den Korpora von Fremdsprachenlernern verfasste Texte gespeichert, welche für die Analyse von häufig auftretenden Problemen hilfreich sind. Aus diesen Texten können für Schüler schwierige Fälle abgeleitet und speziell darauf ausgerichtete Übungen erstellt werden. Auch eine Übung, in der Fehler verbessert werden sollen, ist mit den Texten aus diesem Korpus möglich. Die Schüler könnten bei derartigen Texten auch nach der Akzeptabilität bestimmter Sätze oder Textausschnitte gefragt werden. Die Antworten, welche die Schüler in ihren Tests geben, können wiederum in einem kleinen Lernerkorpus gespeichert und so miteinander verglichen werden. Beim nächsten Test kann auf diese Weise schnell herausgefunden werden, ob sich die Fehler wiederholen oder ob sich die Sprache und die verwendeten Strukturen verbessert haben. Allerdings erfordert die Erstellung eines solchen Korpus einen gewissen Arbeitsaufwand, da die Texte der Schüler, obgleich sie bereits digital vorliegen, zunächst getaggt und eventuell geparst werden müssen. Aus diesem Grund ist es fraglich, ob sich der enorme Aufwand in diesem Fall lohnen würde.

Außerdem ist es möglich, mithilfe eines Abgleichs der von Schülern geschriebenen Texte in einem Lernerkorpus mit einem Referenzkorpus den Umfang des von den einzelnen Schülern beherrschten Wortschatzes zu untersuchen. Um die Bandbreite des Vokabulars eines Schülers herauszufinden, werden die vom Schüler verwendeten Lexeme ermittelt und sowohl quantitativ als auch qualitativ ausgewertet. Die Ergebnisse der Schüler können daraufhin miteinander verglichen werden und verhindern somit falsche Interpretationen der Daten. Es könnten somit mithilfe der Korpora Aussagen über die Kenntnis des Vokabulars getroffen werden. Allerdings muss auch bei dieser Anwendung der Korpora erwähnt werden, dass die Erstellung von Lernerkorpora

¹⁰³⁰ James/ Garrett (1992): op. cit., S. 19.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

zeitaufwändig ist und daher nicht ohne Weiteres durchgeführt werden kann. Es ist also möglich derartige Daten zu erheben, allerdings verbunden mit einem gewissen Zeitaufwand.

Zudem kann mithilfe eines Lernerkorpus das Sprachbewusstsein gefördert werden. Der Lehrer kann den Schülern beispielsweise falsche oder ungünstig formulierte Beispiele aus einem Lernerkorpus zur Verfügung stellen und diese sollen die Fehler verbessern oder die Sätze auf andere Art und Weise formulieren. Solche Übungen sind ebenso in Gruppenarbeit möglich, da die Lernenden mit ihren Mitschülern über die beste Formulierung diskutieren und andere Möglichkeiten gegeneinander abwägen können. Ein Lernerkorpus ist als Quelle für derartige Übungen besser geeignet als erfundene inkorrekte Beispiele, da diese Fehler wirklich von Lernern gemacht wurden und die Schüler dadurch vielleicht ihre eigenen Fehler erkennen und folglich verbessern können. Die Erstellung eines Arbeitsblatts mit Sätzen aus dem Lernerkorpus CORELE (siehe Kapitel 2.2.2.1.) wird in Kapitel 4.2.6. veranschaulicht.

Des Weiteren ist es möglich, anhand eines Lernerkorpus zu ermitteln, in welchen Textarten oder bei welchen Themen die Schüler in Sprachproduktionen die meisten Probleme haben. Mithilfe dieser Erkenntnisse kann der Lehrer in den jeweiligen Klassen, aber auch bei Schulklassen in den folgenden Jahren speziell diesen Problemen vorbeugen. Außerdem können Lerner mithilfe eines Referenzkorpus und eines eigenen Lernerkorpus Vergleiche zwischen den sprachlichen Strukturen der Muttersprachler und der Nichtmuttersprachler durchführen und so ihre eigene Sprachproduktion verbessern.

Für die eigene Erstellung von Korpora gibt es einige Programme, welche kostenlos zur Verfügung stehen. Zwei Beispiele für solche Konkordanzprogramme sind *Microconcord* von Mike Scott und Tim Johns (1993), welches auf *WordSmith*¹⁰³¹ zur Verfügung gestellt wird, und *AntConc*¹⁰³² von Lawrence Anthony, welches auf dessen Internetseite heruntergeladen werden kann. Mit derartigen Programmen können nicht nur Lernerkorpora, sondern jegliche Art von Korpora erstellt werden, wie beispielsweise ein Spezialkorpus zur Bearbeitung und Nutzung der Texte eines bestimmten Themengebiets für den Unterricht. In Kapitel 4.7. wird anhand des Programms *AntConc* veranschaulicht, wie ein Korpus selbst erstellt werden kann.

¹⁰³¹ Vgl. Scott, Mike/ Johns, Tim: *Free Software. Microconcord*. Online Publikation: <http://lexically.net/software/index.htm> (15.07.2017).

¹⁰³² Vgl. Anthony (2018): *AntConc (Version 3.5.7) [Macintosh OS X]*. Op. cit. (14.05.2018).

Egal ob ein Lernerkorpus oder ein selbst erstelltes Spezialkorpus für Unterrichtszwecke, das Korpus, mit dem die Schüler umgehen, sollte laut Osborne nicht zu groß sein, da sie sich ansonsten von der Flut an Daten, welche sie systematisch durchsehen sollen, erschlagen fühlen. Er schreibt in seinem Aufsatz „What can students learn from a corpus?: building bridges between data and explanation“, dass große Korpora zwar repräsentativer seien, die Fülle an Information die Schüler aber überfordern würde und diese daher sowieso nur einen willkürlich ausgewählten Unterbereich analysieren würden.¹⁰³³

Im englischsprachigen Bereich gibt es bereits einige Studien zum Spracherwerb, bei denen Lernerkorpora erstellt wurden. Yukio Tono untersuchte anhand eines Lernerkorpus (des *Japanese English as a Foreign Language Learner Corpus* – JEFLC – mit ungefähr 300.000 Lexemen) die Aneignung des Englischen von japanischen Schülern. Dabei wurde die richtige beziehungsweise falsche Verwendung grammatischer Morpheme, wie des Plural-s, des Genitiv-s, des ‚s‘ der 3. Person Singular Präsens und der unregelmäßigen Formen des *Simple Past*, analysiert. Die Daten der Schüler, welche aus Aufsätzen zu Themen wie *What do you prefer for breakfast, rice or bread?* oder *What would you take out with you in case of a major earthquake?* stammten, wurden nach Wortarten getaggt und im Anschluss wurden die zu untersuchenden grammatischen Morpheme annotiert, um diese herausfiltern und manuell beziehungsweise halbautomatisch untersuchen zu können.¹⁰³⁴

Mithilfe von Lernerkorpora oder von Korpora mit Daten von Fremdsprachenlernern und Muttersprachlern können auch Vergleiche zur Aneignung bestimmter für die jeweilige Sprache typischer Strukturen durchgeführt werden. Anhand des eigens dafür erstellten *Hong Kong Corpus of Spoken English* (HKCSE) wurde eine Studie zur Verwendung von *Question Tags*, *Discourse Markers* (wie *actually*, *really* oder *well*), *Inexplicitness* und *Vague Language* im gesprochenen Englisch von Muttersprachlern und Fremdsprachenlernern in HongKong durchgeführt. Derartige Studien sollen helfen, den Fremdsprachenlernern Englisch so beizubringen, wie Muttersprachler es sprechen. Anhand der Daten aus dem Korpus können die

¹⁰³³ Vgl. Osborne (2000): op. cit., S. 169.

¹⁰³⁴ Vgl. Tono (2000): op. cit., S. 123-132.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

Fremdsprachenlerner auch selbst die Sprache der Muttersprachler analysieren und Schlüsse daraus ziehen in Bezug auf ihren eigenen Sprachgebrauch.¹⁰³⁵

Seidlhofer erstellte mit Studenten aus ihrem Proseminar ein Korpus aus selbst geschriebenen *Summaries* (Zusammenfassungen eines Zeitungsartikels) und *Accounts* (eine Art Reaktion auf diesen Zeitungsartikel) und untersuchte mit ihrem Kurs das Thema Intertextualität anhand dieser Textsammlung. Sie machte den Studenten alle geschriebenen Texte anonymisiert zugänglich und war erfreut, dass diese so interessiert an den Texten der anderen waren. Vor allem ihr Wissen über den Quelltext und die Aufgabenstellung zusammen mit der Verfügbarkeit verschiedener Texte zu diesem Thema führte dazu, dass die Lerner sich in der Lage sahen, den Standpunkt der anderen zu verstehen, aber auch deren Texte zu kritisieren. Dabei stellten die Studenten Fragen zu den Texten, welche anhand des Korpus geklärt werden sollten, beispielsweise, wie viele Lexeme durchschnittlich in den *Summaries* verwendet wurden, welche Zeitformen vorkamen, ob die Lexeme in den selbst geschriebenen Texten dieselben waren wie im Originaltext, ob die fünf häufigsten Begriffe in den *Summaries* mit denjenigen in den *Accounts* übereinstimmten und wie bestimmte Schlüsselwörter verwendet wurden. Außerdem konnte ein Vergleich der *Summaries* mit dem Originaltext Aufschluss über die Wiederholung des ursprünglichen Wortlauts geben, wodurch beibehaltene Strukturen schnell aufgedeckt werden konnten. Zudem stellte sich die Frage, ob es möglich wäre, anhand eines solchen Lernerkorpus eine Art Liste der häufigsten inkorrekten Kollokationen zu erstellen, damit eine Konkordanzübung für die Schüler erarbeitet und somit auf individuelle Fehler reagiert werden könnte. Laut Seidlhofer sollte bei der Verwendung eines Lernerkorpus auch immer ein Referenzkorpus zu Vergleichszwecken herangezogen werden. Anhand dieses Referenzkorpus können das Vorkommen und die Häufigkeit der Schlüsselwörter in muttersprachlichen Texten ermittelt werden und es

¹⁰³⁵ Vgl. Cheng, Winnie/ Warren, Marten (2000): „The HongKong Corpus of Spoken English: language learning through language description“. In: Burnard, L./ McEnery, T. (Hrsg.) (2000): *Rethinking language pedagogy from a corpus perspective: papers from the Third International Conference on Teaching and Language Corpora*. Frankfurt am Main: Lang, S. 133-144.

kann festgestellt werden, ob der Gebrauch bestimmter Begriffe aus den Texten der Schüler mit deren Gebrauch im Referenzkorpus übereinstimmt.¹⁰³⁶

Díaz-Negrillo/Ballier/Thompson (2013) beschäftigen sich mit Design und Annotation von Lernerkorpora, der Automatisierung der Analyse von Lernerdaten sowie der Eignung analytischer Modelle für die Analyse von Lernerdaten.¹⁰³⁷ De Alba Quiñones führte anhand eines schriftlichen Korpus mit Texten deutscher Spanischlerner aus Nordrhein-Westfalen eine Studie zum Gebrauch einfacher lexikalischer Einheiten und Kollokationen durch. Sie verglich richtige und falsche Gebrauchsmuster und erstellte einige Vorschläge zur Vermittlung spanischer Kollokationen im Fremdsprachenunterricht.¹⁰³⁸

Im spanischsprachigen Bereich ist zudem Alonso-Ramos (2016) zu nennen.¹⁰³⁹ Hier finden sich einige Studien zur Erstellung und Nutzung spanischer Lernerkorpora. Es wird besonders betont, dass es sich um eine Aufsatzsammlung handelt, die sich nicht mit dem Englischen als Zweit- oder Fremdsprache beschäftigt, wie dies bei den meisten Werken dieser Art der Fall ist. Damit soll es auf diesem Gebiet die erste Monografie sein, die sich mit einer anderen Sprache als dem Englischen beschäftigt. Alonso-Ramos gibt an, dass die spanischen Lernerkorpora insgesamt zwar 20% der nicht-englischen Lernerkorpora ausmachen, insgesamt allerdings nur einen Anteil von 8% einnehmen, was ihrem Status als einer der am weitesten verbreiteten Sprachen der Welt nicht gerecht wird.¹⁰⁴⁰

Anhand der bereits beschriebenen Korpora SPLLOC (siehe Kapitel 2.2.2.2.) und CEDEL2 (siehe Kapitel 2.2.2.6.) wurde eine Studie zum Gebrauch von Diskursmarkern in der *Interlanguage* von englischen Spanischlernern mittels einer qualitativen und quantitativen Analyse mündlicher und schriftlicher Sprachproben durchgeführt. Es wurden die Interferenz in Bezug auf eine semantische Erweiterung des Gebrauchs von

¹⁰³⁶ Vgl. Seidlhofer, Barbara (2000): „Operationalizing intertextuality: using learner corpora for learning“. In: Burnard, L./ McEnery, T. (Hrsg.) (2000): *Rethinking language pedagogy from a corpus perspective: papers from the Third International Conference on Teaching and Language Corpora*. Frankfurt am Main: Lang, S. 207-223.

¹⁰³⁷ Vgl. Díaz-Negrillo, Ana/ Ballier, Nicolas/ Thompson, Paul (Hrsg.) (2013): *Automatic Treatment and Analysis of Learner Corpus Data*. Amsterdam/ Philadelphia: John Benjamins.

¹⁰³⁸ Vgl. De Alba Quiñones, Virginia (2012): *Lexías simples y colocaciones léxicas en la enseñanza de ELE. Análisis e implicaciones didácticas*. Frankfurt am Main: Peter Lang.

¹⁰³⁹ Vgl. Alonso-Ramos, Margarita (Hrsg.) (2016): *Spanish Learner Corpus Research: Current trends and future perspectives*. Amsterdam [u.a.]: John Benjamins.

¹⁰⁴⁰ Vgl. Alonso-Ramos (2016): op. cit., S. 4.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

oh, actualmente, eventualmente, usualmente und *entonces*, die lexikalisch-pragmatischen Fehler aufgrund semantischer Simplifikation sowie einige andere lexikalisch-pragmatische Fehler hinsichtlich des Gebrauchs von Diskursmarkern untersucht.¹⁰⁴¹ Mithilfe des Fehler-getaggten Teils des Korpus CEDEL2 wurde zudem die Internetseite HARenEs (*Herramienta de Ayuda a la Redacción en Español*) entwickelt, welche Spanischlernern dabei hilft, korrekte Kollokationen in ihren Textproduktionen zu verwenden. Auf dieser Seite können nicht nur Kollokationen gesucht, sondern auch vom Nutzer geschriebene Texte auf die korrekte Verwendung von Kollokationen überprüft sowie ein eigenes Kollokationswörterbuch erstellt werden.¹⁰⁴²

Campillos Llanos untersuchte anhand des mündlichen Lernerkorpus CORELE (siehe Kapitel 2.2.2.1.) den Gebrauch von Artikeln in den Sprachproduktionen von Muttersprachlern und Lernern des Spanischen. Dabei kam er zu dem Ergebnis, dass bestimmte Artikel häufiger verwendet wurden und Lerner, deren Muttersprache über keine Artikel verfügt (wie Polnisch, Chinesisch, Japanisch, Koreanisch, Finnisch und Türkisch), diese Kategorie zu selten nutzten.¹⁰⁴³ Blanco Canales analysierte den Einfluss der drei Variablen Geschlecht, Sprachniveau und Muttersprache auf die phonische Wahrnehmung des Spanischen als Zweitsprache anhand des Korpus *Fono.ele* (siehe Kapitel 2.2.2.3.). Ein Hörverstehenstest führte zu dem Ergebnis, dass der Einfluss dieser Variablen gering ist. Im Gegensatz zur Sprachproduktion, welche ganz klar vom phonetischen System der Muttersprache beeinflusst wird, scheint die Rezeption nicht derart eng mit der Muttersprache in Verbindung zu stehen, da die Ergebnisse des Hörverstehens relativ homogen sind.¹⁰⁴⁴

Des Weiteren konnte anhand der in Kapitel 2.2.2.4. vorgestellten LANGSNAP Lernerkorpora ermittelt werden, dass alle englischen Studenten, die das dritte Jahr ihres

¹⁰⁴¹ Vgl. Vázquez Veiga, Nancy (2016): „Discourse markers in CEDEL2 and SPLLOC corpora of learner Spanish. Analysis of some lexical-pragmatic failures“. In: Alonso-Ramos, M. (Hrsg.) (2016): *Spanish Learner Corpus Research: Current trends and future perspectives*. Amsterdam [u.a.]: John Benjamins, S. 267-297.

¹⁰⁴² Vgl. TALN, DiCE (ed.): *HARenEs. ¡Aprende colocaciones!*. Online Publikation: <http://harenes.taln.upf.edu/CakeHARenEs/> (14.10.2017).

¹⁰⁴³ Vgl. Campillos Llanos, Leonardo (2016): „POS-tagging a Spanish oral learner corpus: Criteria, procedure, and a sample analysis“. In: Alonso-Ramos, M. (Hrsg.) (2016): *Spanish Learner Corpus Research: Current trends and future perspectives*. Amsterdam [u.a.]: John Benjamins, S. 89-116.

¹⁰⁴⁴ Vgl. Blanco Canales, Ana (2016): „Factors that can have an impact on the processes of perceiving Spanish/L2“. In: Alonso-Ramos, M. (Hrsg.) (2016): *Spanish Learner Corpus Research: Current trends and future perspectives*. Amsterdam [u.a.]: John Benjamins, S. 199-234.

vierjährigen Sprachstudiums im Ausland verbrachten, sprachlich davon profitierten, obgleich, wie erwartet, einige Unterschiede zwischen einzelnen Teilnehmern zu verzeichnen waren. Zudem konnten sie ihr Selbstvertrauen und ihre Fähigkeit zur Problemlösung aufbauen. Außerdem wurde festgestellt, dass die englische Sprache in der internationalen Hochschulbildung so präsent ist, dass alle Studenten eine Art bilinguales Leben führten, wodurch es ihnen bisweilen schwer fiel, Spanisch oder Französisch zu sprechen anstatt Englisch.¹⁰⁴⁵

Alonso-Ramos gibt an, dass die Forschung und Ausschöpfung der Möglichkeiten der Lernerkorpora im spanischen Bereich noch weiter vorangebracht werden müssen, wie am Beispiel der englischen Sprache gesehen werden kann. Die Erkenntnisse aus Lernerkorpora werden beispielsweise noch nicht für die Erstellung von Lehrbüchern verwendet. Zudem ist die Nutzung von Lernerdaten durch Schüler selbst noch nicht sehr verbreitet, wenn überhaupt, dann wohl im Bereich der englischen Sprache auf akademischem Niveau. Nach Meinung von Alonso-Ramos muss vor allem die Zusammenarbeit zwischen Verlegern, Lehrern und Wissenschaftlern verbessert werden, um Fortschritte auf diesem Gebiet erzielen zu können. Sie betont, dass ein umfassendes Lernerkorpus nach dem Vorbild des englischen *Cambridge Learner Corpus*, beispielsweise aus den DELE-Tests (*Diploma de Español como Lengua Extranjera*) oder den bereits digitalisierten SIELE-Tests (*Servicio Internacional de Evaluación de la Lengua Española*), wünschenswert wäre, da es nicht nur die Auswahl derjenigen linguistischen Einheiten, die in Lehrbüchern vermittelt werden, sondern auch deren Reihenfolge erleichtern würde. Spanischlehrern liege zwar der *Plan* des *Instituto Cervantes* (1997-2018)¹⁰⁴⁶ vor, die Reihenfolge der Einführung lexikalischer Einheiten scheint ihrer Meinung nach allerdings hauptsächlich auf der Intuition des Autors zu basieren.¹⁰⁴⁷

Eine Möglichkeit der Korpusarbeit für den Spanischunterricht wäre die Erstellung eines Korpus aus Jugendzeitschriften, anhand dessen Jugendsprache untersucht und beispielsweise der Aufbau von Leserbriefen oder ähnlichen Artikeln analysiert werden

¹⁰⁴⁵ Vgl. Mitchell, Rosamond/ Richard, Laurence/ Romero de Mills, Patricia: *The LANGSNAP Project: Languages and Social Networks Abroad*. Op. cit. (12.10.2017).

¹⁰⁴⁶ Vgl. Instituto Cervantes (ed.1997-2018): *Plan curricular del Instituto Cervantes. Niveles de referencia para el español*. Online Publikation: http://cvc.cervantes.es/ensenanza/biblioteca_ele/plan_curricular/ (21.09.2018).

¹⁰⁴⁷ Vgl. Alonso-Ramos (2016): op. cit., S. 16-18.

3. Theoretische Bedingungen, Motive und Möglichkeiten für den Einsatz von Korpora im Fremdsprachenunterricht

könnte, um schließlich selbst einen solchen zu verfassen. Diese Korpuserstellung wäre als Projekt im Sinne des *enfoque por tareas* denkbar (siehe Kapitel 3.1.1.) oder in größerem Umfang auch als wissenschaftliches Seminar (W-Seminar), wie es in der gymnasialen Oberstufe in Bayern von jedem Schüler verpflichtend belegt werden muss.¹⁰⁴⁸ Es handelt sich hierbei allerdings um kein Lernerkorpus im engeren Sinn, sondern vielmehr um ein Spezialkorpus oder ein LSP-Korpus, wie die bereits erwähnten Korpora zur Untersuchung von *Business English*. Die Arbeit mit einem Korpus der Jugendsprache wäre zudem im Sinne des Lehrplans, der für die Qualifikationsphase im Bereich Sprachreflexion angibt, die Schüler sollen „neuere Entwicklungen des Spanischen kennenlernen: Neologismen, regionale und überregionale Sprachentwicklungen, z. B. *Espanglish*, Jugendsprache“. ¹⁰⁴⁹ Auch bei Spanisch als spätbeginnender Fremdsprache ist die Jugendsprache eine Möglichkeit des Kennenlernens spanischer Sprachregister, welche im Lehrplan vorgeschlagen wird.¹⁰⁵⁰ Die Nutzung eines bereits existierenden Korpus der Jugendsprache wird in Kapitel 4.5.2. anhand des Korpus COLA (siehe Kapitel 2.1.2.1.) vorgestellt.

Wie in diesem Kapitel zu sehen war, ist der Gebrauch von Korpora durchaus im Sinne des Lehrplans, des *Gemeinsamen Europäischen Referenzrahmens für Sprachen* sowie aktueller Ansätze des Fremdsprachenunterrichts. Er fördert sogar einige wichtige Prinzipien derselben und unterstützt damit den Lerner in seinem individuellen Lernprozess. Des Weiteren wurde ersichtlich, dass sich die Nutzung von Korpora sowohl für Lehrbuchautoren beziehungsweise generell für die Ersteller didaktischer Materialien als auch für Lehrer und Schüler gewinnbringend auswirkt. Erstere können sie nutzen, um Sprachvielfalt anhand realer Daten aus dem Ursprungsland über ihre Materialien ins Klassenzimmer zu bringen sowie Daten zur Frequenz von Einheiten bei der Erstellung von Wörterbüchern, Grammatiken, Lehrwerken und Materialien miteinzubeziehen. Die Lehrkraft kann das Korpus als reiche und objektive Quelle für Informationen und

¹⁰⁴⁸ Vgl. Bayerisches Staatsministerium für Unterricht und Kultus (2007): *Die Seminare in der gymnasialen Oberstufe*. Online Publikation: https://www.isb.bayern.de/download/1581/isb_seminare_komplett_2-aufl.pdf (01.05.2018).

¹⁰⁴⁹ Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Spanisch (3. Fremdsprache) – Jahrgangsstufe 11/12*. Op. cit. (06.09.2017).

¹⁰⁵⁰ Vgl. Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Spanisch (spätbeginnend) – Jahrgangsstufe 11*. Op. cit. (06.09.2017).

Unterrichtsmaterialien verwenden, welche sowohl ihr selbst bei Zweifeln im Sprachgebrauch hilft als auch Beispielsätze für den Unterricht bereithält. Die Schüler können einerseits von all den eben erwähnten Vorteilen für Autoren und Lehrer profitieren, da deren Bemühungen stets auf eine Verbesserung ihres Lernens abzielen, andererseits ist auch für sie eine direkte Nutzung von Korpora möglich, bei der sie selbstständig im Korpus forschen und ihren eigenen Lernprozess damit steuern. Somit können Korpora den Lernprozess auf vielfältige Art und Weise unterstützen.

Laut McEnery/Xiao werden Korpora bisher vielmehr zur Ermittlung dessen verwendet, was unterrichtet wird, als dafür, wie unterrichtet wird. Das heißt, dass die indirekten Verwendungsmöglichkeiten von Korpora klar überwiegen, während Korpora kaum direkt für den oder im Unterricht genutzt werden. Die direkten Nutzungsmöglichkeiten sind bislang auf die Bereiche der universitären Bildung beschränkt, an weiterbildenden Schulen und im Fremdsprachenunterricht werden korpusbasierte Lernphasen leider kaum durchgeführt.¹⁰⁵¹ Einige Gründe dafür, warum der Gebrauch von Korpora bisher mehr auf indirekte als auf direkte Weise Eingang in den Sprachunterricht findet, sind das Sprachniveau und die Erfahrung der Lerner, der Zeitdruck, die Anforderungen des Lehrplans, das Wissen und die Fähigkeiten der Lehrer in Bezug auf die Korpusanalyse und die Vermittlung des Inhalts an ihre Schüler sowie die Verfügbarkeit der Ausstattung, wie Computer, Software und Korpora.¹⁰⁵² Aus diesem Grund beschäftigt sich die vorliegende Arbeit hauptsächlich mit der direkten Nutzung von Korpora im Unterricht und zeigt Wege und Möglichkeiten, wie diese in den Spanischunterricht integriert werden können.

¹⁰⁵¹ Vgl. McEnery/ Xiao (2010): op. cit., S. 374.

¹⁰⁵² Vgl. McEnery/ Xiao (2010): op. cit., S. 365.

4. Nutzung von Korpora in der fremdsprachlichen Unterrichtspraxis

Korpora können für den Fremdsprachenunterricht am besten im Bereich der Lexik genutzt werden, da vor allem Kollokationen, Synonyme, Phraseologismen oder auch Wortbildungsmuster in Korpora gesucht und für eigene Sprachproduktionen verwendet werden können. Die selbstständige Erarbeitung von Grammatikeinheiten ist im Gegensatz dazu schwieriger, weshalb die Schüler eine genaue Anleitung dafür benötigen, wie sie vorgehen sollen. Dennoch soll in dieser Arbeit veranschaulicht werden, dass Korpora in verschiedenen Bereichen des schulischen Fremdsprachenunterrichts genutzt werden können, so auch für die Aussprache, das Textverständnis sowie landeskundliche Inhalte.

In diesem praxisorientierten Kapitel sollen einige Möglichkeiten präsentiert werden, wie Korpora im und für den Unterricht verwendet werden können. Es kann und soll hierbei kein Anspruch auf Vollständigkeit erhoben werden, da es aufgrund der Fülle an Korpora sowie der vielen unterschiedlichen Nutzungsmöglichkeiten derselben unmöglich ist, alle Optionen jedes Korpus aufzuzeigen. Somit handelt es sich um eine Präsentation verschiedener, für den Unterricht geeigneter und teilweise bereits erprobter Vorgehensweisen, welche zur Nutzung der Korpora sowie zur weiteren Exploration derselben anregen soll.

4.1. Lexik

Da im Bereich des Wortschatzes die Nutzung von Korpora am erfolgversprechendsten ist, handelt es sich hierbei um den umfassendsten Abschnitt dieses Kapitels. Vor allem in Bezug auf die Ermittlung von Kollokationen bieten Korpora Möglichkeiten, die bis dahin kaum vorstellbar waren, doch auch hinsichtlich Phraseologismen, Synonymen, Homonymen oder der Wortbildung können Korpora für den Unterricht genutzt werden.

4.1.1. Kollokationen

Jeremy Clear, ein langjähriger Mitarbeiter im *Cobuild Team* der *University of Birmingham* unter der Leitung von John Sinclair, befasst sich mit der traditionellen Grammatik im Sinne des *Slot-and-filler*-Modells. Diese Sicht der Grammatik als *Open-Choice Principle*, welche sich seit den 1950er Jahren herauskristallisierte, sieht Texte als

eine Reihe von Lücken (*slots*), welche von einem Lexikon, das lokale Restriktionen befriedigt, gefüllt werden müssen. Diesem Prinzip setzt Clear das *Idiom Principle* entgegen, bei welchem dem Sprecher eine große Anzahl an halb vorgefertigten Phrasen zur Verfügung steht, die einzelne Wahlmöglichkeiten darstellen, obwohl es so scheint, als seien sie in Segmente unterteilbar.¹⁰⁵³ Laut John Sinclair dominiert das *Idiom Principle* unser Verständnis von natürlicher Sprache. Wo es allerdings nötig ist, wechseln die Sprecher zum *Open-Choice Principle* und springen dann wieder zurück. Dies geschieht zum Beispiel, wenn eine ungewöhnliche Wortwahl einen solchen Wechsel erforderlich macht. Daher kann nicht auf das *Open-Choice Principle* verzichtet werden und es kann nur eine Kombination aus beiden Prinzipien die Wortwahl erklären.¹⁰⁵⁴

Auch wenn die Existenz der freien Wortwahl nicht vollkommen abgelehnt werden kann und das *Idiom Principle* dem *Open-Choice Principle* bisweilen weichen muss, machen feste Wendungen doch einen beachtlichen Teil jeder Sprache aus. Daher erscheint es durchaus sinnvoll, in den Lehrwerken und im Fremdsprachenunterricht nicht nur einzelne Lexeme zu vermitteln, sondern den Schülern Phraseologismen (siehe Kapitel 4.1.2.) und vor allem Kollokationen beizubringen. Wie viele andere ist allerdings auch der Begriff der Kollokation in der Literatur nicht eindeutig definiert.

Lemnitzer/Zinsmeister definieren den Begriff „Kollokation“ wie folgt:

„Als K. wird das wiederholte gemeinsame Vorkommen zweier Wörter in einer strukturell interessanten Einheit bezeichnet. In einer Kollokation beeinflusst ein Wort die Auswahl eines anderen Wortes zuungunsten von Wörtern mit gleicher oder ähnlicher Bedeutung.“¹⁰⁵⁵

Dabei betonen die Autoren, dass dieses Kovorkommen nicht zufällig sein darf. Ein lexikalisches Zeichen muss ein anderes vielmehr als dessen sprachlichen Kontext bestimmen, meist unter Ausschluss anderer, bedeutungsähnlicher Zeichen.¹⁰⁵⁶ Diese Bestimmung des Kontexts sieht auch Michael Hoey als Kernelement der Definition und bezeichnet sie als die lexikalischen Präferenzen eines Lexems.¹⁰⁵⁷ Lemnitzer/Zinsmeister

¹⁰⁵³ Vgl. Clear, Jeremy (2000): „Do you believe in Grammar?“. In: Burnard, L./ McEnery, T. (Hrsg.) (2000): *Rethinking language pedagogy from a corpus perspective: papers from the Third International Conference on Teaching and Language Corpora*. Frankfurt am Main: Peter Lang, S. 19-25.

¹⁰⁵⁴ Vgl. Sinclair, John McHardy (1987): „Collocation: a progress report“. In Steele, R./ Threadgold, T. (Hrsg.) (1987): *Language topics: essays in honour of Michael Halliday*. Amsterdam: John Benjamins, S. 324.

¹⁰⁵⁵ Lemnitzer/ Zinsmeister (2010): op. cit., S. 188.

¹⁰⁵⁶ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 142.

¹⁰⁵⁷ Vgl. Hoey, Michael (2000): „The hidden lexical clues of textual organisation: a preliminary investigation into an unusual text from corpus perspective“. In Burnard, L./ McEnery, T. (Hrsg.) (2000):

halten zudem fest, dass die Auswahl eines Lexems durch ein anderes arbiträr und auf die jeweilige Einzelsprache beschränkt ist. Es handelt sich daher um komplexe sprachliche Zeichen. Des Weiteren lässt sich diese Auswahl nicht regelhaft beschreiben, weshalb die Wortverbindungen als Ganzes gelernt werden müssen.¹⁰⁵⁸

Lemnitzer/Zinsmeister unterscheiden zudem genau zwischen den Begriffen *Kookkurrenz* und *Kollokation*. Ersterer bezeichnet das gemeinsame Vorkommen von zwei Lexemen in einem gemeinsamen sprachlichen Kontext, wobei die Reihenfolge zweier Lexeme als unterscheidendes Kriterium zweier Kookkurrenzen dient. Sofern zwei Einheiten in einem Textfenster häufiger zusammen vorkommen als bei einer zufälligen Verteilung, spricht man von signifikantem Kovorkommen.¹⁰⁵⁹ Scherer hingegen gibt an, dass der Begriff *Kollokation* eben dieses gemeinsame Vorkommen zweier Lexeme in einem mehr oder weniger eng umgrenzten sprachlichen Kontext beschreibt. Die Tatsache, dass das Vorkommen deutlich die Wahrscheinlichkeit eines zufälligen Zusammentreffens übersteigt, wird bei ihr als *signifikante Kollokation* bezeichnet. Sie setzt den Begriff *Kollokation* also mit demjenigen der *Kookkurrenz* gleich und benennt damit zwei oder mehrere überdurchschnittlich oft benachbarte Lexeme.¹⁰⁶⁰ Im Sinne des bloßen gemeinsamen Vorkommens in einem gemeinsamen Kontext wurde der Begriff „Kollokation“ erstmals von Firth (1957)¹⁰⁶¹ im Rahmen des Kontextualismus aufgebracht, er wurde dann aber eingegrenzt auf diejenigen Paare von Lexemen, die üblicherweise miteinander vorkommen. Greenbaum unterscheidet im Jahr 1970 zwischen dem *item-oriented approach*, welcher lediglich auf linearen Kookkurrenzen basiert, und dem *integrated approach*, der syntaktische und semantische Relationen berücksichtigt.¹⁰⁶²

„A more valuable, if more modest, contribution might be made to the study of collocations if a relatively homogenous class of items were selected and an investigation undertaken of the collocation of each item in the class with other items that are related syntactically in a given way.“¹⁰⁶³

Rethinking Language Pedagogy from a Corpus Perspective. Frankfurt am Main: Peter Lang, S. 35.

¹⁰⁵⁸ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 142.

¹⁰⁵⁹ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 143-144.

¹⁰⁶⁰ Vgl. Scherer (2006): op. cit., S. 46-47.

¹⁰⁶¹ Vgl. Firth, John Rupert (1957): *Papers in Linguistics 1934-1951*. London [u.a.]: Oxford University Press.

¹⁰⁶² Vgl. Greenbaum, Sidney (1970): *Verb-Intensifier collocations in English. An experimental approach*. The Hague: Mouton, S. 10-14.

¹⁰⁶³ Greenbaum (1970): op. cit., S. 13.

Nach dieser Eingrenzung Greenbaums können Verbindungen zwischen Substantiven und Adjektiven oder Substantiven und Verben als Kollokationen angesehen werden, zufällig zusammen vorkommende lexikalische Einheiten ohne syntaktische Beziehung jedoch nicht. Hausmann (1984, 1985) bringt den Begriff mit Kombinationsbeschränkungen in Verbindung und führt schließlich die Unterscheidung zwischen Basis und Kollokator ein. Zwischen ihnen herrscht eine gerichtete Beziehung - die Basis bestimmt den Kollokator. Während die Basis semantisch autonom ist und auch mit anderen Lexemen frei kombiniert werden kann, ist die Verwendung des Kollokators stark restriktiv.¹⁰⁶⁴ Pöll gibt ebenso an, dass die Basis semantisch autonom und daher unproblematisch ist, der Kollokator allerdings Probleme bereiten kann, da er idiosynkratisch und somit nicht vorhersehbar ist.¹⁰⁶⁵ Er kann häufig fast ausschließlich mit dieser Basis verwendet werden. Hausmann führt hier als Beispiel die Kollokation *schütteres Haar* an, bei deren Verwendung der Textproduzent nicht etwa das Lexem *schütter* verwenden möchte und händeringend nach dem Substantiv *Haar* sucht, sondern vielmehr über das Haar einer Person schreiben möchte und dabei als passendes Adjektiv *schütter* auswählt.¹⁰⁶⁶ In Bezug auf dieses Adjektiv spricht er ebenfalls die Unterschiede im Kollokationspotential eines Lexems an. So kann das Adjektiv *schütter* fast ausschließlich mit dem Substantiv *Haar* verwendet werden und hat daher nur ein sehr begrenztes Kollokationspotential. Das Lexem *kurz* hat, im Gegensatz dazu, aufgrund der vielen Kombinationsmöglichkeiten ein sehr weites Kollokationspotential.¹⁰⁶⁷

Auch Lemnitzer/Zinsmeister gemäß muss eine Kollokation sowohl die Kriterien der Kookkurrenz erfüllen als auch eine innere Struktur in Form einer Hierarchie zwischen Kollokationsbasis und Kollokator aufweisen. Außerdem müssen die lexikalischen Einheiten bei einer Kollokation in einer syntaktischen Beziehung zueinander stehen, zum Beispiel als Köpfe einer Verbalphrase und einer gleich- oder untergeordneten Nominalphrase, wie in *hacer cola* (‚Schlange stehen‘), oder als Kopf einer

¹⁰⁶⁴ Vgl. Hausmann, Franz Josef (1984): „Wortschatzlernen ist Kollokationslernen. Zum Lehren und Lernen französischer Wortverbindungen“. In: *Praxis des neusprachlichen Unterrichts* 1984, Bd. 31, S. 401-402.

¹⁰⁶⁵ Vgl. Pöll, Bernhard (2002): *Spanische Lexikologie. Eine Einführung*. Tübingen: Narr, S. 86.

¹⁰⁶⁶ Vgl. Hausmann (1984): op. cit., S. 401-402.

¹⁰⁶⁷ Vgl. Hausmann, Franz Josef (1985): „Kollokationen im deutschen Wörterbuch. Ein Beitrag zur Theorie des lexikographischen Beispiels“. In: Bergenholtz, H./ Mugdan, J. (Hrsg.) (1985): *Lexikographie und Grammatik. Akten des Essener Kolloquiums zur Grammatik im Wörterbuch 28.-30.06.1984*. Tübingen: Niemeyer, S. 124.

Nominalphrase und Kopf einer untergeordneten Adjektivphrase.¹⁰⁶⁸ Das Substantiv, als wichtigste Basiswortart, welche die Welt versprachlicht, stellt bei Kollokationen meist die Basis dar und bildet mit Verben und Adjektiven eine Kollokation. Letztere stellen dann das Basiswort einer Kollokation dar, wenn sie durch Adverbien weiter determiniert werden.¹⁰⁶⁹

In Bußmanns *Lexikon der Sprachwissenschaft* werden Kollokationen definiert als “charakteristische, häufig auftretende Wortverbindungen, deren gemeinsames Vorkommen auf Regelmäßigkeit gegenseitiger Erwartbarkeit beruht, also primär semantisch (nicht grammatisch) begründet ist”.¹⁰⁷⁰ Hier kommt also zusätzlich die Frequenz des Auftretens ins Spiel. Auch Scherer zufolge kann die Stärke einer Kollokation anhand der Häufigkeit des gemeinsamen Auftretens bestimmt werden.¹⁰⁷¹

Mellado Blanco ist allerdings der Meinung, dass in Bezug auf die Sprache im Allgemeinen, sofern es sich also nicht um *lenguajes de especialidad* handelt, die bloße statistische Konzeption der Kollokation zu reduktionistisch scheint, da es frequente Kombinationen gibt, die eben nicht als Kollokationen angesehen werden können. Für Fachsprachen ist eine statistische Herangehensweise ihrer Meinung nach durchaus von Bedeutung, da Fachbegriffe unter anderem anhand der Frequenz ermittelt werden, doch generell sind Kollokationen ein syntaktisch-semantisches Phänomen. Die Häufigkeit spielt daher nur eine untergeordnete Rolle.¹⁰⁷²

Im Hinblick darauf erwähnt Hausmann (2004) eine Art Krieg zwischen dem basisbezogenen Kollokationsbegriff, wie er im Fremdsprachenunterricht und der Lexikographie verwendet wird, und dem computerlinguistischen, der lediglich auf die Häufigkeit der Verbindungen Wert legt, also nicht auf *to book a hotel/ to check into a hotel*, sondern auf jegliche sogenannte *cluster* wie *hotel at, hotel room, a hotel, either hotel, left the hotel*.¹⁰⁷³ Letzterer ist vielmehr mit Lemnitzer/Zinsmeisters Begriff der

¹⁰⁶⁸ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 144.

¹⁰⁶⁹ Vgl. Hausmann (1985): op. cit., S. 119.

¹⁰⁷⁰ Bußmann (2008): op. cit., S. 345.

¹⁰⁷¹ Vgl. Scherer (2006): op. cit., S. 46-47.

¹⁰⁷² Vgl. Mellado Blanco, Carmen (2008): „Introducción: colocaciones y algunas cuestiones teórico-prácticas de fraseografía“. In: Mellado Blanco, C. (2008): *Colocaciones y fraseología en los diccionarios*. Frankfurt am Main: Peter Lang, S. 15.

¹⁰⁷³ Vgl. Hausmann, Franz Josef (2004): „Was sind eigentlich Kollokationen“. In: Steyer, K. (2004) (Hrsg.): *Wortverbindungen – mehr oder weniger fest*. Berlin: De Gruyter, S. 319-322.

Kookkurrenz gleichzusetzen.¹⁰⁷⁴ Auch für Irsula Peña spielt vor allem die Tatsache, dass es sich um typische, relevante und gebräuchliche Kombinationen handelt, eine wichtige Rolle. Er schreibt dem Erlernen von Kollokationen im Fremdsprachenunterricht besonderen Wert zu und definiert sie wie folgt:¹⁰⁷⁵

„Kollokationen sind sachgerechte, grammatisch einwandfreie, einzelsprachlich typische, relevante, spezifische, usuelle, stabile und somit reproduzierbare Kombinationen, die entsprechende typische Sachverhalte, Objekte und Erscheinungen der Realität bezeichnen.“¹⁰⁷⁶

Bustos Plaza unterscheidet zwei verschiedene Kollokationsbegriffe: einen statistischen und einen semantischen. Der statistische Kollokationsbegriff ist vor allem in der Anglistik vorherrschend und bezeichnet „häufige, syntaktisch zusammengehörige Wortfolgen“.¹⁰⁷⁷ In diesem Sinne gibt Hunston beispielsweise an: „Collocation is the statistical tendency of words to co-occur.“¹⁰⁷⁸ Die Häufigkeit spielt nach Meinung von Bustos Plaza allerdings nicht die bedeutendste Rolle bei der Kategorisierung von Kollokationen. Der semantische Kollokationsbegriff, der in den oben erwähnten Arbeiten von Hausmann seinen Anfang fand, beschreibt Kollokationen als „stabile, von den Sprechern bekannte und reproduzierbare Zweierkombinationen“, welche „in der Normebene und nicht im Sprachsystem angesiedelt“ sind (siehe auch Kapitel 1.1.1.). Das Sprachsystem ließe an sich auch andere Kombinationen zu. Innerhalb dieser Auffassung spielt nicht das Kriterium der Häufigkeit, sondern das der Verfügbarkeit eine wichtige Rolle.¹⁰⁷⁹ Pöll unterstreicht ebenso die Tatsache, dass es sich, entgegen der weit verbreiteten Ansicht, bei Kollokationen keineswegs um besonders häufige Verbindungen handelt. Dass die Frequenz vielerseits als Definitionskriterium gesehen wird, liegt seiner Meinung nach an den quantitativen Studien in der Anglistik.¹⁰⁸⁰

Laut Hausmanns Charakterisierung des Begriffs Kollokation führen frequenzbasierte Analysen zu keinem brauchbaren Ergebnis, da seiner Meinung nach unspezifische Verbindungen wie *Angst haben* viel häufiger verwendet würden als

¹⁰⁷⁴ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 143-144.

¹⁰⁷⁵ Vgl. Irsula Peña, Jesús (1994): *Substantiv-Verb-Kollokationen: Kontrastive Untersuchungen Deutsch-Spanisch*. Frankfurt am Main: Peter Lang, S. 140.

¹⁰⁷⁶ Irsula Peña (1994): op. cit., S. 50.

¹⁰⁷⁷ Bustos Plaza, Alberto (2012): „Kollokationen und Funktionsverbgefüge“. In: Born, J./ Folger, R./ Laferl, C. F./ Pöll, B. (Hrsg.) (2012): *Handbuch Spanisch*. Berlin: Schmidt, S. 253.

¹⁰⁷⁸ Hunston (2002): op. cit., S. 12.

¹⁰⁷⁹ Bustos Plaza (2012): op. cit., S. 252-253.

¹⁰⁸⁰ Vgl. Pöll (2002): op. cit., S. 84.

spezifische, in der Kombinatorik eingeschränkte, wie beispielsweise *Angst einflößen* oder *Angst einjagen*. Derartige Kollokationen, die zwar verfügbar oder disponibel, jedoch nicht frequent sind, können folglich nicht durch Frequenzanalysen als Kollokationen belegt werden.¹⁰⁸¹

Weiter bringt Hausmann den Begriff der Kollokationsspanne (*collocational span*) aus dem britischen Kontextualismus ins Gespräch. Kollokationspartner stehen nicht immer direkt nebeneinander wie in *schütteres Haar*, sondern können durch ein weiteres Lexem getrennt sein, wie in *Das Haar ist schütter*, oder sogar durch den Großteil des Satzes, wie in *Das Haar ist nicht nur bei alten Menschen, sondern auch bei relativ jungen Menschen bereits recht häufig schütter*.¹⁰⁸² Sinclair bestimmt die gewöhnliche Nähe mit maximal vier dazwischen liegenden Lexemen.¹⁰⁸³

Da die Kollokationspartner also nicht immer direkt nebeneinander stehen müssen, ist es folglich schwer, sie mittels Frequenzanalysen zu ermitteln. Zudem sind Frequenzanalysen laut Pöll bei denjenigen Kollokationen sinnlos, welche die einzige Möglichkeit der Versprachlichung eines Sachverhalts darstellen, wie beispielsweise beim Ausdruck *asumir una responsabilidad*, für den es kaum strukturell mehrgliedrige Alternativen gibt. Sie sind allerdings interessant, wenn mehrere äquivalente Kollokationen zum Ausdruck eines Sachverhalts zur Verfügung stehen. Als Beispiele werden hier *anular/ cancelar/ rescindir un contrato*, *dar/ atribuir importancia a* und *echar/ pronunciar un discurso* genannt.¹⁰⁸⁴

Für Benson gilt es, lexikalische von grammatikalischen Kollokationen zu unterscheiden. Eine lexikalische Kollokation besteht seiner Meinung nach aus zwei ‚gleichwertigen‘ lexikalischen Komponenten wie Adjektiv – Substantiv, Substantiv – Verb oder Verb – Substantiv, welche gewöhnlich zusammen auftreten, wie *tomar una decisión*. Diese Art der Kollokation wird meist in Lehrbüchern vermittelt. Im Gegensatz dazu besteht eine grammatikalische Kollokation aus einem ‚dominanten‘ Lexem, also einem Substantiv, Adjektiv oder Verb, und einem darauffolgenden ‚grammatikalischen‘ Lexem wie einer Präposition. Spanische Beispiele wären demnach *consistir en* oder

¹⁰⁸¹ Vgl. Hausmann (1985): op. cit., S. 124-126.

¹⁰⁸² Vgl. Hausmann (1985): op. cit., S. 127.

¹⁰⁸³ Vgl. Sinclair (1991): op. cit., S. 170.

¹⁰⁸⁴ Vgl. Pöll (2002): op. cit., S. 84.

acceso a.¹⁰⁸⁵ Aufgrund der Tatsache, dass die eine Art von Kollokationen syntaktisch ist und die andere lexikalisch, können beide Typen auch zusammenfallen. Ein Beispiel hierfür ist *caer en la trampa*, mit der grammatikalischen Kollokation *caer en* und der lexikalischen Kollokation *caer en – la trampa*.¹⁰⁸⁶ Koike geht ebenfalls von dieser Unterscheidung Bensons zwischen lexikalischen und grammatikalischen Kollokationen aus und wählt für seine Untersuchungen lexikalische Kollokationen des Typs Substantiv–Verb und Substantiv–Adjektiv aus, da diese hinsichtlich der Anzahl und der kommunikativen Bedeutung die repräsentativsten sind.¹⁰⁸⁷

Laut Iliescu sind Kollokationen „Wortsequenzen, die relativ stabil sind und mehr oder weniger feste Syntagmen bilden“.¹⁰⁸⁸ Im Gegensatz zu Hausmann (1984)¹⁰⁸⁹ sieht sie ihre Definition als viel breiter an, da sie nicht mehr von ‚fixierten‘ Wortbildungen spricht. Bei Kollokationen geht es ihrer Meinung nach um Wortsequenzen, die gewissen, für einen Nicht-Muttersprachler unvorhersehbaren Restriktionen unterliegen. Nach Iliescu existieren zwei verschiedene Arten von Kollokationen. Auf der einen Seite die konvergierenden Kollokationen, welche sich ihrer Meinung nach auf einer höheren sprachlichen Ebene befinden und der gemeinsamen europäischen althergebrachten Kultur zu verdanken sind, auf der anderen Seite die divergierenden Kollokationen, die aus der gesprochenen Umgangssprache stammen, meist neuerer Herkunft und eher idiosynkratisch sind. Allerdings spiegeln letztere auch die zeitgenössischen wirtschaftlichen, sozialen und politischen Umstände wider und tendieren daher heute im vereinten Europa und in der globalisierten Welt in vielen Fällen ebenfalls zur Konvergenz.¹⁰⁹⁰ Hausmann schreibt die Kollokation als „typische, spezifische und charakteristische Zweierkombination von Wörtern“ im Hinblick auf Saussures Dichotomie der *langue* zu. „Kollokationen des Typs *eingefleischter Junggeselle, schütteres Haar, heikles Thema* [...] sind wenn nicht Fertigprodukte, so wenigstens

¹⁰⁸⁵ Vgl. Benson, Morton (1985): „Collocation and idioms“. In: Ilson, R. (Hrsg.) (1985): *Dictionaries, lexicography and language learning*. Oxford: British Council/ Pergamon, S. 61-62.

¹⁰⁸⁶ Vgl. Koike, Kazumi (2001): *Colocaciones léxicas en el español actual: Estudio formal y léxico-semántico*. Alcalá de Henares: Universidad de Alcalá de Henares [u.a.], S. 14.

¹⁰⁸⁷ Vgl. Koike (2001): op. cit., S. 14.

¹⁰⁸⁸ Iliescu, Maria (2006): „Kollokationen in den romanischen Sprachen“. In: Dietrich, W./ Hoinkes, U./ Roviró, B./ Warnecke, M. (Hrsg.) (2006): *Lexikalische Semantik und Korpuslinguistik*. Tübingen: Narr Francke Attempto Verlag, S. 190.

¹⁰⁸⁹ Vgl. Hausmann (1984): op. cit., S. 398.

¹⁰⁹⁰ Vgl. Iliescu (2006): op. cit., S. 189-190.

4. Nutzung von Korpora in der fremdsprachlichen Unterrichtspraxis

Halbfertigprodukte der Sprache, zwar nicht der Sprache als System, aber im Sinne Coserius der Sprache als Norm“ (siehe Kapitel 1.1.1.).¹⁰⁹¹ Diese müssen seiner Meinung nach von unspezifischen banalen Kombinationen, wie *Lineal benutzen* oder *kleiner Koffer* unterschieden werden, welche nicht Teil eines Wörterbuchs sein sollten.¹⁰⁹²

Von Bedeutung ist dabei, Kollokationen einerseits von den eben angesprochenen freien Verbindungen abzugrenzen, andererseits von Phraseologismen und Wortbildungsprodukten, mit denen sie den Status von im mentalen Lexikon verankerten Einheiten teilen.¹⁰⁹³ Pöll verwendet für diese Abgrenzung ein von Hausmann für das Französische entwickelte Schema und bezieht es auf das Spanische:

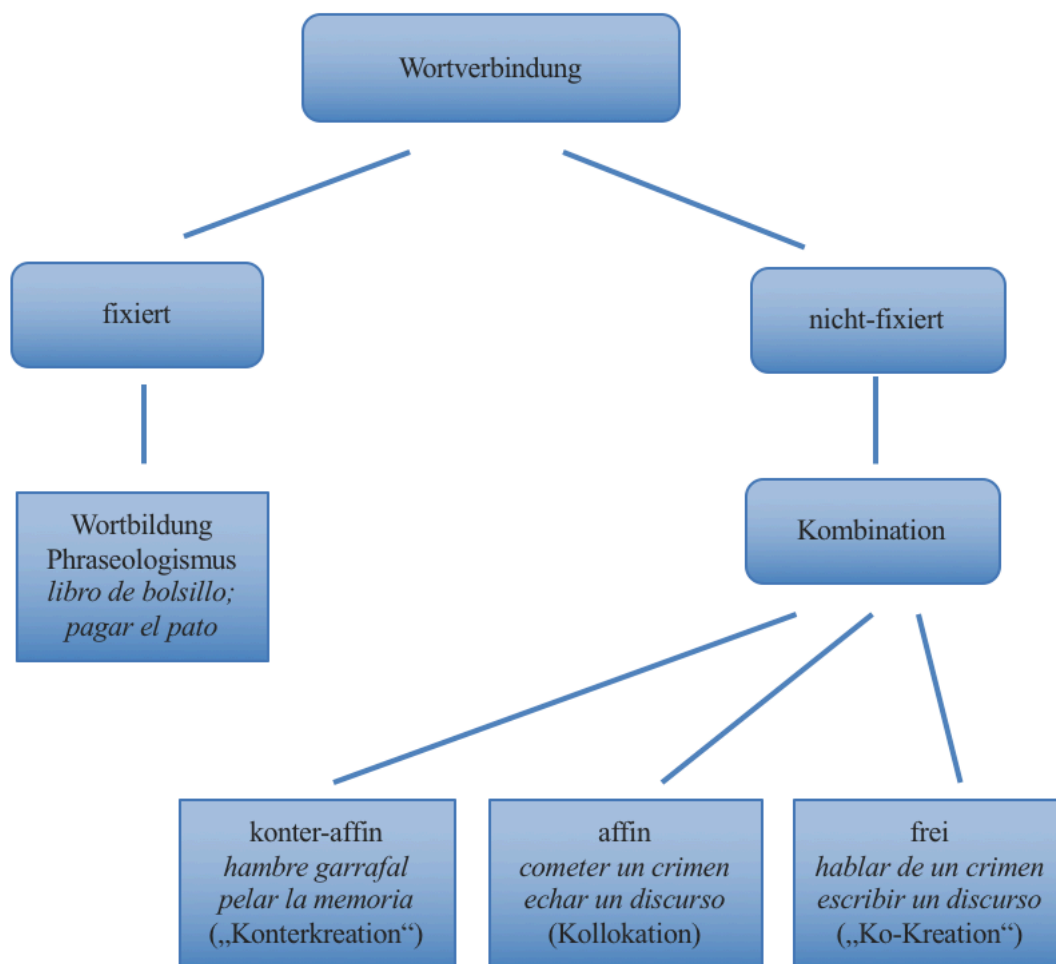


Abbildung 79: Schema zur Abgrenzung von Kollokationen nach Pöll (2002).¹⁰⁹⁴

¹⁰⁹¹ Hausmann (1985): op. cit., S. 118.

¹⁰⁹² Vgl. Hausmann (1985): op. cit., S. 120-121.

¹⁰⁹³ Vgl. Pöll (2002): op. cit., S. 84-85.

¹⁰⁹⁴ Vgl. Pöll (2002): op. cit., S. 85. In Anlehnung an: Hausmann (1984): op. cit., S. 399.

Den Kollokationen werden seitens der fixierten Wortverbindungen Phraseologismen und komplexe Lexeme gegenübergestellt. Letztere wurden nach syntagmatischen Prinzipien gebildet und zeichnen sich wie Phraseologismen durch die innere Kohärenz aus. Unter den nicht-fixierten Kombinationen gilt es, Kollokationen von Konterkreationen und Ko-Kreationen abzugrenzen. Während Konterkreationen nicht-lexikalisierte Verbindungen darstellen, bei denen semantische Beschränkungen verletzt werden, stellen Ko-Kreationen freie Verbindungen dar, die den erwartbaren semantischen Beschränkungen unterliegen.¹⁰⁹⁵ Hausmann charakterisiert die von ihm im obigen Schema benannten Gruppen wie folgt: „Unter dem Gesichtspunkt der Üblichkeit sind die Kollokationen Kombinationen von auffälliger Üblichkeit, die Ko-Kreationen solche von unauffälliger Üblichkeit und die Konter-Kreationen solche von auffälliger Unüblichkeit.“¹⁰⁹⁶

Weiter gibt Pöll an, Kollokationen unterscheiden sich von Phraseologismen dadurch, dass sie anders als Letztere Transformationen zulassen, beispielsweise durch Einschub eines Adjektivs wie bei *correr un gran peligro*.¹⁰⁹⁷ Für Benson liegt der Unterschied zwischen Kollokationen und Phraseologismen, welche er als *idioms* bezeichnet, darin, dass die Bedeutung der Ersteren mehr oder weniger aus den Bedeutungen der einzelnen Teile abgeleitet werden können, obgleich die Präpositionen in Kollokationen nicht vorherzusehen sind.¹⁰⁹⁸ Der Unterschied zu Wortbildungsprodukten ist derjenige, dass diese der Begriffsbildung dienen, wohingegen Kollokationen komplexe Situationen versprachlichen und verschiedene Konzepte zueinander in Beziehung setzen, beispielsweise bei *crimen sexual* vs. *crimen atroz*. Auf der anderen Seite muss man sie von freien Verbindungen abgrenzen. Dabei ist zu beachten, dass Kollokationen semantisch reguläre Verbindungen sind, deren Besonderheit eben darin besteht, dass andere semantisch ähnliche Verbindungen nicht mit der gleichen Basis zusammengesetzt werden. Beispiele sind *diente de ajo* und nicht **dedo de ajo*, *levantar/ verter/ difundir/ propagar calumnias* und nicht **levantar/ verter/ difundir/ propagar falsedades*.¹⁰⁹⁹ Die Beschränkung der Kombination von Kollokationen kann häufig anhand eines Sprachvergleichs verdeutlicht werden. Beispiele

¹⁰⁹⁵ Vgl. Pöll (2002): op. cit., S. 85.

¹⁰⁹⁶ Hausmann (1984): op. cit., S. 399.

¹⁰⁹⁷ Vgl. Pöll (2002): op. cit., S. 86.

¹⁰⁹⁸ Vgl. Benson (1985): op. cit., S. 62.

¹⁰⁹⁹ Vgl. Pöll (2002): op. cit., S. 87.

hierfür sind *acariciar* (wörtlich: ‚liebkosten‘) *una idea* im Gegensatz zum deutschen ‚mit einer Idee spielen‘, *diente* (‚Zahn‘) *de ajo* gegenüber der deutschen ‚Knoblauchzehe‘ oder *tomar* (‚nehmen‘) *una decision* mit der deutschen Entsprechung ‚eine Entscheidung treffen‘. Im Vergleich zu anderen Sprachen, vor allem den romanischen, finden sich aber auch ähnliche Kombinationsbeschränkungen für diese Konzepte, beispielsweise das spanische *tomar una decision* und das französische *prendre une décision*, spanisch *diente de ajo* und portugiesisch *dente de alho* sowie spanisch *acariciar una idea*, französisch *caresser une idée* und portugiesisch *afagar uma ideia*.¹¹⁰⁰

Koike gibt an, dass es drei formale und drei semantische Kriterien für die Charakterisierung lexikalischer Kollokationen gibt. Zu den formalen Merkmalen zählen häufiges Kovorkommen, kombinatorische Restriktionen sowie formale Kompositionalität. Zunächst ist also das häufige gemeinsame Vorkommen zweier lexikalischer Einheiten eines der wichtigsten Kennzeichen der Kollokation, obgleich es kein Alleinstellungsmerkmal ist, da nicht alle häufig vorkommenden Kombinationen lexikalische Kollokationen sind und umgekehrt auch nicht alle Kollokationen sehr frequent sind. Auch grammatikalische Kollokationen und phraseologische Einheiten zeichnen sich durch einen hohen Grad der Kookkurrenz aus. Zudem müssen die beiden lexikalischen Einheiten nicht immer direkt nebeneinander stehen, sondern lediglich innerhalb der bereits genannten Kollokationsspanne. Zweitens sind sie bevorzugte oder gewohnheitsmäßige Kombinationen, die aufgrund des traditionellen Gebrauchs bestimmten kombinatorischen Restriktionen unterliegen. Diese sorgen dafür, dass eine Einheit häufig zusammen mit einer bestimmten anderen unter Ausschluss der Auswahl anderer, gleichbedeutender Einheiten auftritt. Drittens unterliegen sie einer gewissen formalen Kompositionalität, sie erlauben also die Substitution einer Komponente (beispielsweise *violar/ transgredir/ vulnerar las normas*), Modifizierung durch Adjektive (beispielsweise *daba una explicación satisfactoria*), Pronominalisierung, Nominalisierung (beispielsweise *repicar la campana – repique de campanas*), Transformation ins Passiv und dergleichen.¹¹⁰¹

In semantischer Hinsicht charakterisieren sich Kollokationen durch die Bindung, die zwei Lexeme miteinander eingehen, die Typizität der Beziehung, die zwischen diesen

¹¹⁰⁰ Vgl. Pöll (2002): op. cit., S. 87.

¹¹⁰¹ Vgl. Koike (2001): op. cit., S. 25-28.

beiden Lexemen besteht, sowie die semantische Präzision, die durch sie repräsentiert wird. Da die Kollokation in der Regel eine Bindung zweier lexikalischer Einheiten darstellt, ist es möglich, dass einige von ihnen die Wortart wechseln, beispielsweise bei *luchar encarnizadamente* – *lucha encarnizada*, *acalorarse [la discusión]* – *discutir acaloradamente* – *discusión acalorada*. Des Weiteren drücken sie eine typische Beziehung zwischen ihren Komponenten aus, im Fall einer Substantiv-Verb-Kollokation eben zwischen dem Substantiv und dem Verb. Aufgrund der Typizität der Beziehung, die zwischen den Komponenten besteht, sind *tocar la guitarra* und *rasguear la guitarra* Kollokationen, *limpiar la guitarra* allerdings nicht, da hier keine typische Beziehung zu einem Musikinstrument ausgedrückt wird. Zuletzt ist die semantische Präzision zu nennen. Im Gegensatz zu Phraseologismen, welche bei Koike als *locuciones* bezeichnet werden, präsentieren Kollokationen, insbesondere des Typs Substantiv–Verb, eine semantische Präzision, aufgrund derer sie dafür eingesetzt werden, um einfache lexikalische Einheiten in Wörterbucheinträgen zu definieren. So lässt sich in der ersten Bedeutung von *rumor* die Kollokation *correr {noticia}* (*noticia vaga u oficiosa que corre entre la gente*) finden. In der Bedeutung von *rumorear* (*circular un rumor, transmitirse de unos a otros*) findet sich die Kollokation *circular {rumor}*.¹¹⁰²

Mittels dieser sechs Charakteristika lassen sich Kollokationen einerseits von freien Kombinationen und andererseits von Phraseologismen abgrenzen. Von freien Kombinationen unterscheiden sich Kollokationen dadurch, dass Erstere nicht generell durch eine Häufigkeit des Kovorkommens gekennzeichnet sind. Die Kookkurrenz liegt hier vielmehr im Ermessen des Sprechers. Zudem zeichnen sich freie Kombinationen durch einen geringeren Grad an kombinatorischen Restriktionen sowie eine größere kombinatorische, morphologische und syntaktische Flexibilität aus. Des Weiteren drücken sie keine typische Beziehung zwischen zwei Einheiten aus, wie am Unterschied zwischen *tocar la guitarra* und *limpiar la guitarra* gesehen werden kann.¹¹⁰³

Im Gegensatz zu Phraseologismen sind Kollokationen in syntaktischer Hinsicht wiederum durch eine größere formale Flexibilität gekennzeichnet. Erstere erlauben keinerlei Modifizierung durch Adjektive (mit Ausnahme der Quantifikation oder Intensifikation) oder Pronominalisierung des Substantivs in der Funktion des direkten

¹¹⁰² Vgl. Koike (2001): op. cit., S. 28-29.

¹¹⁰³ Vgl. Koike (2001): op. cit., S. 30.

Objekts (beispielsweise in: *Asumió el cargo de alcalde, pero su repentina enfermedad le impidió desempeñarlo*). Außerdem kann das Substantiv eines Phraseologismus nicht das Bezugswort eines Relativsatzes sein, wie im Fall der Kollokation. Des Weiteren erlauben sie weder Transformation ins Passiv noch Nominalisierung, wie anhand der Nominalisierung der Kollokation *trasplantar un órgano* zu *el trasplante del órgano* im Gegensatz zum Phraseologismus *escurrir el bulto* (**el escurrimiento del bulto*) zu sehen ist. Im Hinblick auf die Semantik enthalten Phraseologismen eine größere Anzahl an implizierten Lexemen, während eine Kollokation meist aus zwei Einheiten besteht. Zudem sind sie semantisch weniger transparent als Kollokationen, welche meist durch eine semantische Kompositionalität gekennzeichnet sind. Phraseologismen haben außerdem zwei Bedeutungen, eine wortwörtliche und eine übertragene, wobei Erstere eine geringe Rolle spielt. Schließlich zeichnen sich Phraseologismen nicht durch eine typische Beziehung zwischen den Einheiten aus, wie an *pagar el pato* im Gegensatz zur Kollokation *tocar el piano* ersichtlich wird. Trotz dieser detaillierten Aufzählung an Unterschieden muss Koike allerdings zugeben, dass es einige Kombinationen gibt, die sowohl als Kollokationen als auch als Phraseologismen gelten können.¹¹⁰⁴

Bustos Plaza schreibt andererseits Funktionsverbgefüge des Typs Verb-Substantiv (beispielsweise *poner en movimiento*), welche er definiert als „verbäquivalente Kombinationen [...], die aus einem grammatikalisierten, weitgehend desemantisierten Verbalelement und einem Nominalement bestehen, das seiner Semantik nach eher verbaler Natur ist“, den Kollokationen zu, da sich eine hypothetische Grammatikalisierung des Verbs nicht rechtfertigen lässt.¹¹⁰⁵

Es gibt also unterschiedliche Definitionen, Ein- und Zuteilungen des Terminus Kollokation, welche von einer engen Definition des Begriffs Phraseologie, der die Kollokation aufgrund des Mangels an Idiomatizität ausschließt, über eine weite, sie einschließende Definition desselben bis hin zur Kategorisierung der Funktionsverbgefüge als Kollokationen führt. Die Einteilung der Kollokation als Teil der Phraseologie wirft Fragen auf bezüglich des Grads der Lexikalisierung des Ausdrucks, der Grenzen der Austauschbarkeit lexikalischer Komponenten durch andere Äquivalente, des Grads der Desemantisierung des Verbs sowie der extralinguistischen Prädetermination, welche die

¹¹⁰⁴ Vgl. Koike (2001): op. cit., S. 31-37.

¹¹⁰⁵ Bustos Plaza (2012): op. cit., S. 252.

Basis auf den verbalen Kollokator ausübt. Die Zuweisung der Funktionsverbgefüge zu den Kollokationen wird ihrerseits von einigen Sprachwissenschaftlern, wie beispielsweise Rivas González, kritisiert, da diese Konstruktionen nicht als Kollokationen gesehen werden können, sofern das Hilfsverb delexikalisiert ist und nunmehr lediglich die Funktion eines Morphems hat, es sich also nicht um zwei referentielle Komponenten handelt. Auf der anderen Seite vertreten nicht nur Bustos Plaza, sondern auch Koike und Corpas Pastor eine weite Definition der Kollokation, welche die Funktionsverbgefüge miteinschließt.¹¹⁰⁶ Wie so oft, herrscht auch in Bezug auf den Begriff der Kollokation keine Einigkeit darüber, welche Kombinationen damit gemeint sind und welche nicht.

In der vorliegenden Arbeit soll unter dem Begriff ‚Kollokation‘ im Sinne der lexikalischen Kollokation nach Benson (1986) eine (mehr oder weniger häufig vorkommende) Verbindung zweier lexikalischer Einheiten verstanden werden, die sich durch eine syntaktische Beziehung untereinander sowie durch formale Flexibilität, semantische Kompositionalität und eine gewisse Typizität der Beziehung auszeichnet.

Es wird also von einem weiten Kollokationsbegriff ausgegangen, der sich von Kookkurrenzen durch die syntaktische Beziehung zwischen den Komponenten unterscheidet, von den freien Kombinationen durch die Häufigkeit des Kovorkommens und von den Phraseologismen durch die formale Flexibilität, die semantische Kompositionalität sowie die Typizität der Beziehung. Während somit Kookkurrenzen des Typs *él viene* oder *y después* und freie Kombinationen wie *pelo lindo* oder *casa bonita* ausgeschlossen werden, können beispielsweise in das Kollokationsfeld *pelo* nicht nur die zur Beschreibung des Haars typischen Adjektive wie *rubio* und *rizado* aufgenommen werden, sondern eben auch *negro*, obwohl es sich hierbei nicht um eine Kombinationsbeschränkung im Sinne Hausmanns (1984, 1985)¹¹⁰⁷ handelt, da dieses Adjektiv nicht ausschließlich zur Beschreibung des Haars verwendet werden kann. Andererseits werden Phraseologismen komplett ausgeschlossen aufgrund der Unvorhersehbarkeit ihrer Bedeutung, welche sich eben nicht aus den Einzelbedeutungen der Bestandteile ergibt (für eine genaue Definition des Begriffs Phraseologismus siehe Kapitel 4.1.2.). Im Gegensatz zu Koike (2001), welcher lediglich die Verbindungen

¹¹⁰⁶ Vgl. Mellado Blanco (2008): op. cit., S. 9.

¹¹⁰⁷ Vgl. Hausmann (1984): op. cit., S. 395-406. Und: Hausmann (1985): op. cit., S. 118-129.

Substantiv–Verb, Verb–Substantiv und Substantiv–Adjektiv in seine Studie aufnimmt, sollen hier wie bei Montoro del Arco (2010) auch die Kombinationen Verb–Adverb und Adverb–Adjektiv aufgenommen werden.

Im Hinblick auf den Unterricht gibt Hausmann an, dass es besonders wichtig ist, den Schülern Kollokationen zu vermitteln, da diese Kombinationsbeschränkungen unterliegen. Während Ko-Kreationen ohne besondere Schwierigkeiten zu erlernen sind, da sie beliebig kombiniert werden können und so keine Fehlerquelle darstellen, ist die kreative Kombination von Lexemen im Bereich der Kollokationen nicht möglich, weshalb keine einzelnen Lexeme, sondern Wortverbindungen gelernt werden müssen, um Fehler zu vermeiden. Diese Problematik spielt im rezeptiven Bereich meist keine Rolle, im produktiven dafür umso mehr.¹¹⁰⁸ Laut Montoro del Arco stellt die Beherrschung der Kollokationen einer Sprache einen signifikanten Hinweis auf die kommunikativen Fähigkeiten eines Sprechers dar. Aus diesem Grund misst er dem Erlernen von Kollokationen sowohl bei Muttersprachlern als auch bei Fremdsprachenlernern so große Bedeutung bei.¹¹⁰⁹

Den Schülern sollte schon so früh wie möglich die Brisanz der Verwendung kollokativer Verbindungen deutlich gemacht werden. Oft mangelt es selbst Muttersprachlern in gewissen Punkten an Wissen in diesem Bereich. Laut Stubbs (1995) stellt die Intuition der Muttersprachler nicht unbedingt eine verlässliche Quelle für linguistische Daten dar, vor allem im Hinblick auf Kollokationen. Er stellt klar, dass Muttersprachler zwar zentrale Beispiele von Kollokationen angeben könnten, es jedoch weniger wahrscheinlich sei, von ihnen verlässliche Auskunft über deren Frequenz oder deren Verteilung in unterschiedlichen Registern zu erhalten.¹¹¹⁰ Muttersprachler können laut Buckingham zweifelsfrei angeben, dass das Substantiv *impulso* zusammen mit dem Verb *dar* eine Kollokation bildet, es ist allerdings fraglich, ob sie wissen, dass diese Kollokation im Bereich der Sozialwissenschaften fast immer mit einem modifizierenden Adjektiv verwendet wird und dass die häufigsten Modifikatoren, ihren Daten zufolge, die

¹¹⁰⁸ Vgl. Hausmann (1984): op. cit., S. 400-402.

¹¹⁰⁹ Vgl. Montoro del Arco, Esteban Tomás (2010): „Norma y uso de las colocaciones léxicas del español“. In: Aleza Izquierdo, M. (2010): *Normas y usos correctos en el español actual*. Valencia: Tirant Humanidades, S. 254.

¹¹¹⁰ Vgl. Stubbs, Michael (1995): „Collocations and semantic profiles. On the cause of the trouble with quantitative studies“. In: *Functions of Language* 2, 1 (1995). Online Publikation: <https://www.univ-trier.de/fileadmin/fb2/ANG/Linguistik/Stubbs/stubbs-1995-cause-trouble.pdf> (30.09.2017), S. 2.

folgenden sind: *nuevo, fuerte, renovado, gran, moderado, inicial, principal, oficial, significativo*.¹¹¹¹

Mellado Blanco führt aus, dass es für einen Nicht-Muttersprachler eine Utopie ist, die unendlich vielen Wortkombinationen im Kopf zu haben, welche ein Muttersprachler seit seiner Kindheit verinnerlicht hat und ohne Anstrengung, fast automatisch, nutzt. Die Verwendung dieser lexikalischen Verbindungen ist es jedoch, welche dazu führt, dass ein Muttersprachler im Gespräch mit einem anderen nicht auffällt und als solcher akzeptiert wird. Für Lerner der spanischen Sprache ist diese lexikalische Kombinationslehre jedoch eben nicht vorherzusehen und durch Regeln bestimmbar, sie kann lediglich nachträglich entdeckt, beschrieben und erklärt werden.¹¹¹²

Auch Leńko-Szymańska (2003) gibt an, dass die häufigsten Fehler in Aufsätzen von Fremdsprachenlernern in Zusammenhang mit Kollokationen stehen. Das kontextuelle Verhalten lexikalischer Einheiten ist für Schüler einer der problematischsten Aspekte des Wortschatzes der Zweit- oder Fremdsprache. Dies sei schon des Öfteren in der Literatur zum Zweitspracherwerb und zum Fremdsprachenunterricht besprochen, aber zuvor noch nie systematisch überprüft worden. Obwohl ihre Studie gezeigt hat, dass sich das Wissen um kollokationale Zusammenhänge verbessert, je länger der Schüler die Fremdsprache lernt, ist dieser Bereich doch einer derjenigen, in denen die meisten Fehler gemacht werden. Ein Problem beim Erlernen von Kollokationen ist die Tatsache, dass es in jeder Sprache beachtlich viele vorgefertigte lexikalische Phrasen gibt. Dies führt dazu, dass der Lerner einer Fremdsprache eine äußerst hohe Anzahl an lexikalischen Phrasen erlernen muss, um die Sprache zu beherrschen, die jeweilige Phrase aber, im Gesamtbild betrachtet, nur relativ selten erscheint. Diese Tatsache macht das Erlernen dieser typischen Einheiten der Fremdsprache so schwierig.¹¹¹³

Bei den Schülern des zweiten Lernjahrs zählt Leńko-Szymańska je 10000 Lexeme 131 Fehler, wobei 16 davon Fehler in Bezug auf Kollokationen sind. Es handelt sich also um 12,2% aller Fehler. Im vierten Lernjahr zählt sie noch 57 Fehler je 10000 Lexeme, wobei es sich bei 9 Fehlern, also bei 15,79%, um Kollokationsfehler handelt. Die Anzahl der Fehler sinkt zwar vom zweiten zum vierten Jahr ab, der Anteil der Kollokationsfehler steigt aber von 12,2 auf 15,79% an. Während andere Fehlertypen stark abnehmen, bleiben

¹¹¹¹ Vgl. Buckingham (2009): op. cit., S. 45.

¹¹¹² Vgl. Mellado Blanco (2008): op. cit., S. 8.

¹¹¹³ Vgl. Leńko-Szymańska (2003): op. cit., S. 515-516.

die Fehler aufgrund falscher kollokationaler Kombinationen bestehen. Bei anderen Arten der insgesamt 13 Fehlertypen, die beispielsweise die Rechtschreibung, die Morphologie, den Stil oder das Register betreffen, verhält es sich anders. Die Anzahl der Rechtschreibfehler sinkt vom zweiten zum vierten Lernjahr von 33 auf 9 Fehler je 10000 Lexeme. Morphologiefehler fallen von 7 auf 2 je 10000 Lexeme und die gezählten 11 Fehler im Stil sind im vierten Lernjahr ganz verschwunden.¹¹¹⁴

4.1.1.1. Kollokationen in Wörterbüchern

Hausmann sieht ein großes Problem bei der Textproduktion darin, dass die Wörterbücher Kollokationen unter dem Kollokator auflisten, da dieser ohne die Basis gar nicht definiert werden kann. Auf diese Art und Weise können die Benutzer der Wörterbücher die Möglichkeiten der Verwendung der Basis allerdings nicht ganz ausschöpfen, da sie das Wörterbuch systematisch danach absuchen müssten.¹¹¹⁵ Da derartige Wörterbücher nur für den rezeptiven Bereich geeignet sind, schlägt Hausmann vor, dass Wörterbücher, die sowohl von Textrezipienten als auch von Textproduzenten verwendet werden, Kollokationen sowohl unter dem Eintrag des Kollokators als auch dem der Basis auflisten sollten. Da eine solche doppelte Aufführung von Kollokationen die Nachschlagewerke jedoch um einiges vergrößern würde, wäre auch eine Ergänzung traditioneller, auf Rezeption ausgelegter Wörterbücher durch Kollokationswörterbücher für die Textproduktion denkbar.¹¹¹⁶ Auch Corpas Pastor (1992) beschäftigt sich mit der Auflistung von Kollokationen in Wörterbüchern. Neben der Tatsache, dass die zweisprachigen englisch-spanischen Nachschlagewerke durchaus einige Kollokationen auflisten, kommt sie zu dem Schluss, dass einsprachige englische Wörterbücher nicht nur zahlenmäßig mehr Kollokationen beinhalten als die zweisprachigen, sondern diese zudem unter dem Kollokator und der Basis aufgeführt werden. Ihrer Meinung nach muss auf dem Gebiet der Kollokationen mehr geforscht werden, um so deren Kategorisierung und Einordnung in Wörterbüchern voranzubringen.¹¹¹⁷

¹¹¹⁴ Vgl. Leńko-Szymańska (2003): op. cit., S. 513.

¹¹¹⁵ Vgl. Hausmann (1984): op. cit., S. 401-402.

¹¹¹⁶ Vgl. Hausmann (1985): op. cit., S. 121-123.

¹¹¹⁷ Vgl. Corpas Pastor, Gloria (1992): „Tratamiento de las colocaciones del tipo A+S/S+A en diccionarios bilingües y monolingües (español-inglés)“. In: Alvar Ezquerro, M. (Hrsg.) (1992): *EURALEX '90 Proceedings. Actas del IV Congreso Internacional*. Barcelona: Bibliograf, S. 331-340.

Auch 20 Jahre später bedauert Hausmann weiterhin, dass Wörterbücher Kollokationen unter dem Eintrag des Kollokators auflisten, da es gerade die Basen sind, die den Schülern parat stehen und eben die Kollokatoren dazu gefunden werden müssen. Daher propagiert er einmal mehr die Erstellung von Kollokationswörterbüchern, bei denen Kollokationen nach deren Basis aufgelistet sind, um den Lernern der fremden Sprache die Textproduktion zu erleichtern. Während den Schülern bei der Textproduktion die Basis-Substantive meist bereitstehen, fällt es ihnen oft schwer, die treffende, in jeder Einzelsprache unterschiedliche Kombination aus Substantiv und Verb oder Adjektiv auszuwählen.¹¹¹⁸ Luque Toro (2006) stellt bei seiner Analyse der Auflistung von Kollokationen in spanischen Wörterbüchern ebenso fest, dass diese, wenn sie überhaupt integriert wurden, nur sehr schwer aufzufinden sind. Auch er schlägt daher vor, Kollokationen unter der Basis aufzuführen.¹¹¹⁹

Im Bereich der englischen Sprache wurde im Jahr 2002 das *Oxford Collocations Dictionary for Students of English* veröffentlicht, welches Kollokationen nach dem basisbezogenen Begriff gemäß Hausmann definiert und sie unter der Basis auflistet.¹¹²⁰ Es handelt sich hierbei um ein korpusbasiertes Wörterbuch, das speziell für die Sprachproduktion designt wurde.¹¹²¹ Für das Spanische können hier die bereits in Kapitel 3.3.2.1. vorgestellten Wörterbücher REDES und PRÁCTICO von Ignacio Bosque angeführt werden. Es handelt sich hierbei zwar nicht explizit um Kollokationswörterbücher, aber um *diccionarios combinatorios*, welche die Lexeme nicht definieren, sondern gewöhnliche Kombinationen auflisten. Mithilfe des PRÁCTICO, welches 400.000 gewöhnliche Wortverbindungen enthält, kann gemäß des Erstellers der Wortschatz erweitert und die eigene Ausdrucksweise flüssiger und präziser gestaltet werden.¹¹²² Es wird jeweils angegeben, mit welchen anderen Wortarten der Ausdruck verwendet wird, und die üblichsten Beispiele dafür werden aufgelistet. Im PRÁCTICO werden Kollokationen tatsächlich sowohl unter dem Kollokator als auch unter der Basis aufgeführt. Allerdings enthält es, wie gesagt, nicht nur Kollokationen,

¹¹¹⁸ Vgl. Hausmann (2004): op. cit., S. 309-312.

¹¹¹⁹ Vgl. Luque Toro, Luis (2006): „El concepto de colocación y su presencia en los diccionarios de uso del español actual“. In: Alonso Ramos, M. (Hrsg.) (2006): *Diccionarios y Fraseología*. A Coruña: Servizo de Publ., Univ. da Coruña, S. 127-128.

¹¹²⁰ Vgl. Deuter, Margaret (2002): *Oxford Collocations Dictionary for Students of English*. Oxford: Oxford University Press.

¹¹²¹ Vgl. Deuter (2002): op. cit., S. v-xi.

¹¹²² Vgl. Grupo SM (ed.): op. cit. (03.10.2017).

sondern generell häufige Kombinationen, weshalb der Nutzer unter Umständen einen sehr langen Eintrag durchforsten muss, bis er zur gesuchten Kombination gelangt. Bei der Kollokation *medida drástica* findet man die Basis *medida* als 48. Vertreter in einer Reihe von 83 Substantiven im Eintrag von *drástico/-a* und diesen Kollokator auf Platz 35 unter 65 Adjektiven im Eintrag von *medida*. Im Fall der Kollokation *tocar la guitarra* ist die Basis *guitarra* als zweites Substantiv innerhalb der zweiten Bedeutung von *tocar* angegeben, doch der Kollokator *tocar* findet sich nicht direkt im Eintrag von *guitarra*. Hier wird auf den übergeordneten Eintrag *Instrumento Musical* verwiesen, welcher Kombinationen enthält, die für einen Großteil der Musikinstrumente zutreffen, und in welchem *tocar* an dritter Stelle der Verben steht. Des Weiteren steht im PRÁCTICO das Adjektiv *garrafal* an vierter Stelle des Eintrags von *error* und das Adjektiv *craso* an zwölfter Stelle. Bei *enemigo* ist *acérrimo* ebenso an zwölfter Stelle zu finden.¹¹²³ Mithilfe dieses Wörterbuchs wäre die Erarbeitung von Kollokationen somit durchaus möglich.

In diesem Bereich ist auch das DiCE (*Diccionario de colocaciones del Español*) zu nennen, ein Online-Kollokationswörterbuch, das im Internet kostenlos genutzt werden kann (<http://www.dicesp.com/paginas>). Wie in den beiden Wörterbüchern von Ignacio Bosque werden auch hier die Lexeme nicht definiert, sondern gewöhnliche Kombinationen mit anderen Lexemen gezeigt. Derzeit ist es allerdings auf den semantischen Bereich der Substantive der Gefühle begrenzt.¹¹²⁴ Das DiCE besteht aus zwei Komponenten, einem Wörterbuch (*consulta general*) und der *consulta avanzada*. Für jeden Eintrag werden die Bedeutung, die Argumentstruktur, Beispiele, Synonyme, Antonyme, Kollokationen sowie semantische Derivate angegeben.¹¹²⁵ Bei der *consulta general* kann ein Lemma eingegeben oder ausgewählt werden, um zum Eintrag des jeweiligen Substantivs zu gelangen.¹¹²⁶ Die *consulta avanzada* erlaubt im Vergleich dazu die Suche nach konkreten Kombinationen (Bereich *¿Qué significa?*), die Suche mittels der Basis und der Wortart des Kollokators (Bereich *¿Cómo se dice?*), mittels des Kollokators (Bereich *Búsqueda por colocativo*) sowie mittels der lexikalischen Funktion

¹¹²³ Vgl. Bosque (2006): op. cit., S. XVII.

¹¹²⁴ Vgl. Alonso Ramos, Margarita: *Diccionario de colocaciones del Español*. Online Publikation: <http://www.dicesp.com/paginas> (20.10.2017).

¹¹²⁵ Vgl. Alonso Ramos, Margarita: *Diccionario de colocaciones del Español*. *¿Cómo es el DiCE?*. Online Publikation: <http://www.dicesp.com/paginas/index/1> (20.10.2017).

¹¹²⁶ Vgl. Alonso Ramos, Margarita: *Diccionario de colocaciones del Español*. *Consulta general*. Online Publikation: <http://www.dicesp.com/consultageneral/lemas> (20.10.2017).

(Bereich *Búsqueda por función léxica*). Letztere ermöglicht beispielsweise die Suche nach dem Gegenteil eines Ausdrucks (*Contr(sonrisas)=lágrimas*), einem passenden Adverb (*Adv1(andar)=a pie*) oder einer Steigerung (*Magn(error)=craso*).¹¹²⁷ Damit stellt das DiCE ein nützliches Hilfsmittel für die Arbeit mit Kollokationen dar, wobei es derzeit wegen der Beschränkung auf die Gefühlswelt noch nicht umfänglich genutzt werden kann.

Neben diesen Wörterbüchern können auch auf der bereits in Kapitel 3.3.3. erwähnten Internetseite HARenEs (*Herramienta de Ayuda a la Redacción en Español*) Kollokationen gesucht werden. Bei diesem Programm, welches unter anderem von den Entwicklern des DiCE mithilfe des Fehler-getaggten Teils des Korpus CEDEL2 (siehe Kapitel 2.2.2.6.) erstellt wurde, handelt es sich um ein korpusbasiertes Hilfsmittel für die Suche und Korrektur von Kollokationen.¹¹²⁸ Es ist ein Programm, das ähnlich wie ein Online-Wörterbuch aufgebaut ist, in das aber auch Texte hochgeladen und dort überprüft werden können. Damit ist HARenEs ein verfügbares, einfach anzuwendendes und sogar korpusbasiertes Programm, um Kollokationen zu ermitteln. Doch obwohl auf der Internetseite sehr gut nach den zuvor genannten Basen gesucht werden kann, in deren Einträgen die jeweiligen Kollokatoren auch allesamt zu finden sind, ist die umgekehrte Suche nach den Kollokatoren, unter deren Eintrag sich die jeweilige Basis befinden sollte, nicht möglich. Es erscheinen bei HARenEs beispielsweise sowohl *craso* als auch *garrafal* als Kollokatoren im Eintrag von *error*, doch es findet sich kein Eintrag zu *craso* oder *garrafal* in Kombination mit Substantiven.

Daneben stehen den Schülern bei Hausaufgaben, aber auch in Prüfungssituationen einige Wörterbücher zur Verfügung, welche sie somit nutzen könnten, um gängige Kollokationen in ihre Textproduktionen zu integrieren. Doch auch hier stellt sich die Frage, ob und wie diese Nachschlagewerke Kollokationen auflisten. Unter den gängigen Wörterbüchern, die an bayerischen Gymnasien zu Prüfungszwecken verwendet werden dürfen, finden sich unter anderem das *Diccionario Salamanca de la lengua española*, das

¹¹²⁷ Vgl. Alonso Ramos, Margarita: *Diccionario de colocaciones del Español. Consulta avanzada*. Online Publikation: <http://www.dicesp.com/paginas/index/6> (20.10.2017).

¹¹²⁸ Vgl. TALN, DiCE (ed.): *HARenEs. ¡Aprende colocaciones!*. Op. cit. (14.10.2017).

Langenscheidt Diccionario de Español para Extranjeros sowie das *Cornelsen Klausurwörterbuch Spanisch*.¹¹²⁹

Im einsprachigen Wörterbuch *Diccionario Salamanca de la lengua española* (2006) findet sich *drástico* nicht im Eintrag von *medida*. Im Eintrag von *drástico* wird *medida* zwar nicht speziell aufgelistet, es findet sich jedoch ein Beispielsatz mit dieser Kollokation. Das Verb *tocar* wird nicht speziell unter *guitarra* aufgeführt, doch auch hier findet sich ein Beispiel mit dieser Verbindung. Im Gegensatz zu *piano* und *violón* wird *guitarra* zwar nicht eigens im Eintrag von *tocar* erwähnt, doch auch hier findet sich ein Beispiel mit dieser Kombination. Unter *error* werden weder *garrafal* noch *craso* erwähnt. Unter *craso* wird *error* hingegen durchaus aufgelistet und unter *garrafal* findet sich ein Beispielsatz. Auch *acérrimo* wird bei *enemigo* nicht erwähnt und Letzteres ebenso wenig im Eintrag des Adjektivs. Folglich sind Kollokationen im *Diccionario Salamanca* nicht unter der Basis, sondern unter dem Kollokator aufgeführt. Lediglich einmal findet sich bei einer Basis ein Beispielsatz mit der entsprechenden Kollokation.¹¹³⁰

Im einsprachigen *Langenscheidt sm Diccionario de Español para Extranjeros* (2006) ist die Situation ähnlich. Das Adjektiv *drástico* findet sich nicht im Eintrag von *medida*, unter Ersterem wird die Kollokation jedoch schon geführt. Unter *guitarra* findet sich kein Hinweis auf *tocar*, unter Letzterem wird lediglich die Bedeutung in Kombination mit Musikinstrumenten im Allgemeinen aufgelistet. Unter *error* werden *garrafal* und *craso* wiederum nicht erwähnt, wobei die Kollokation *craso error* durchaus im Eintrag von *craso* zu finden ist, ebenso wie ein Beispielsatz bei *garrafal*. Ebenso wenig wird *acérrimo* bei *enemigo* erwähnt, im Eintrag des Kollokators findet sich allerdings durchaus das Beispiel *enemigos acérrimos*.¹¹³¹

Auch im zweisprachigen *Cornelsen Klausurwörterbuch Spanisch* (2012) ist das Adjektiv *drástico* nicht bei den Angaben zu *medida* zu finden, wohingegen Letzteres

¹¹²⁹ Vgl. Bayerisches Staatsministerium für Bildung und Kultus, Wissenschaft und Kunst (2017): *Für Prüfungszwecke genehmigte Wörterbücher im Bereich der modernen Fremdsprachen*. Online Publikation: https://www.google.de/url?sa=t&rct=j&q=&esrc=s&source=web&cd=3&cad=rja&uact=8&ved=0ahUKEwiOjcuM4PnWAhXjCsAKHQxZAFgQFgg7MAI&url=https%3A%2F%2Fwww.km.bayern.de%2Fdownload%2F11637_liste_genehmigte_wrterbcher_2017_juni.pdf&usg=AOvVaw3enpt99ooGUaerF7LlfZMD (18.10.2017).

¹¹³⁰ Vgl. Gutiérrez Cuadrado, Juan (2006): *Diccionario Salamanca de la lengua española*. Madrid: Santillana [u.a.].

¹¹³¹ Vgl. Arellano, Jesús (2006): *Langenscheidt sm Diccionario de Español para Extranjeros*. München [u.a.]: Langenscheidt.

gleich zu Beginn des Eintrags von *drástico* aufgelistet ist. Unter *guitarra* ist ebenso wenig ein Hinweis auf *tocar* zu finden, wobei im Falle eines zweisprachigen Wörterbuchs die Möglichkeit besteht, unter dem deutschen Verb ‚spielen‘ nachzusehen. Dort sind jedoch lediglich Klavier und Geige als Musikinstrumente aufgeführt. Unter *tocar* findet sich gleich an dritter Stelle die Bedeutung ‚Gitarre spielen‘. Im Eintrag von *error* ist weder ein Verweis auf *craso* noch auf *garrafal* zu finden. Unter *craso* wird *error* auch nicht erwähnt und zu *garrafal* gibt es gar keinen Eintrag im Klausurwörterbuch. Bei *enemigo* findet sich auch kein Hinweis auf *acérrimo* und unter Letzterem wird die Basis ebenso wenig aufgeführt.¹¹³² Zudem werden in zweisprachigen Wörterbüchern meist keine sprachlichen Kontexte für die betreffenden Lexeme angegeben. Eine bloße Aneinanderreihung von Ausdrücken erschwert den Schülern jedoch die Nutzung des Vokabulars im Kontext, da sie häufig nicht wissen, wie die Lexeme in Sätze eingebettet werden.

Wie zu sehen ist, eignen sich die gängigen spanischen Schulwörterbücher im Hinblick auf Kollokationen ebenfalls hauptsächlich für die Rezeption, wie von Hausmann für das Englische beanstandet. Um Kollokationen anhand der Basis zu ermitteln und sie für die Textproduktion zu verwenden, sind sie weniger von Nutzen. Sofern eine Kollokation in der Muttersprache ähnlich gebildet wird, beispielsweise bei *tocar la guitarra* und ‚Gitarre spielen‘, kann der Kollokator unter Umständen mittels eines zweisprachigen Wörterbuchs ermittelt werden. Wenn die Zusammenstellung der Kollokation allerdings nicht aufgrund der Analogie zur eigenen Muttersprache ersichtlich ist, kann auch diese Ressource nicht weiterhelfen.

Derzeit werden außerdem häufig Online-Wörterbücher für Hausaufgaben verwendet aufgrund der Tatsache, dass die meisten Schüler ohnehin einen Computer besitzen und sie somit innerhalb weniger Sekunden das entsprechende Lexem parat haben. Im PONS Online-Wörterbuch findet sich unter dem Substantiv *medida* allerdings kein Hinweis auf den Kollokator *drástico* ebenso wenig wie unter Letzterem die Basis aufgeführt ist. Im recht übersichtlichen Eintrag von *guitarra* ist der Kollokator *tocar* auch nicht erwähnt, es werden lediglich zwei Phraseologismen (*chafar la guitarra a alguien*, *venir como una guitarra en un entierro*) aufgelistet (siehe Kapitel 4.1.2.). Die zweite Bedeutung von *tocar* bei PONS ist zwar ‚spielen‘ im Sinne von *hacer sonar*, es wird

¹¹³² Vgl. Worsch, Wolfgang (2012): *Cornelsen Klausurwörterbuch Spanisch*. Berlin: Cornelsen.

allerdings lediglich die Kombination *tocar la flauta* angegeben. Außerdem findet man weder *garrafal* noch *craso* im Eintrag von *error*. Unter *craso* und *garrafal* wird *error* hingegen durchaus angegeben.¹¹³³ Es ist also auch nicht möglich, diese Kollokationen mittels des PONS Online-Wörterbuchs zu ermitteln. Da Kollokationen demnach häufig nicht in den zu schulischen Zwecken verwendeten Wörterbüchern aufgelistet oder aber lediglich unter dem Kollokator zu finden sind, wo der Schüler sie bei der Suche nach Ausdrücken für die Textproduktion nicht finden kann, sollte im Fremdsprachenunterricht mehr für die Kenntnis von Kollokationen getan werden.

Grundsätzlich sollte sich die Auswahl der Kollokationen zum einen an der Häufigkeit derselben im Sprachgebrauch der Muttersprachler, also beispielsweise an Korpusanalysen, orientieren, so wie eben generell die Auswahl des zu erlernenden Vokabulars, zum anderen aber auch an den semantischen Feldern, welche die Einbettung des Vokabulars in das Lehrbuch erlauben. Hier sollte so vorgegangen werden wie beim übrigen Wortschatz. Sofern ein Lexem aufgrund dessen Häufigkeit, der Nützlichkeit für das jeweilige semantische Feld oder aus anderen, pädagogischen Gründen in den Vokabelteil des Lehrwerks integriert werden soll, könnte auch eine passende Kollokation mit diesem Lexem eingeführt werden. Teilweise wird bereits auf diese Weise vorgegangen, denn bei der Einführung des Substantivs *guitarra* wird meist das Verb *tocar* mit angegeben. Ebenso könnten bei der Basis *error* der Kollokator *garrafal* und bei *medida* das Adjektiv *drástica* angeführt werden.

Aufgrund dieser Schwierigkeiten bei der Verwendung von Kollokationen schlägt Hausmann eine systematische Kollokationsarbeit im Sprachlernprozess vor, die aus vier Grundpfeilern besteht: Der erste Pfeiler ist die semasiologische Kollokationsarbeit. Hier soll der Schüler ausgehend vom Kollokator die mit diesem verbundenen Basen erlernen, um seine rezeptive wie auch seine produktive Kompetenz zu verbessern. Der zweite Pfeiler ist die onomasiologische Kollokationsarbeit, bei der ausgehend von der Basis ein onomasiologisches Kollokationsfeld für die Verbesserung der produktiven Sprachkompetenz erstellt werden soll.¹¹³⁴ Laut Doyé besteht ein Kollokationsfeld „aus all den Wörtern, die nach den Konventionen einer Sprache bevorzugt mit einem

¹¹³³ Vgl. Pons (ed.): *Online Wörterbuch Deutsch – Spanisch*. Online Publikation: <http://de.pons.com> (03.11.2017).

¹¹³⁴ Vgl. Hausmann (1984): op. cit., S. 403-404.

gegebenen (Ausgangs-)Wort zusammen auftreten“.¹¹³⁵ Des Weiteren rät Hausmann zur zweisprachigen Kollokationsarbeit, da zwar Kollokationsbasen in verschiedenen Sprachen äquivalent sind, Kollokatoren jedoch nicht. Letztere werden erst innerhalb der Kollokation zu Äquivalenten in der Mutter- und der Fremdsprache. Die Schüler müssen lernen, dass Kollokatoren nicht wie frei kombinierbare Lexeme einzeln übersetzt werden können, sondern nur als ganze Kollokation. Schließlich verbindet Hausmann die onomasiologische Kollokationsarbeit mit dem Erlernen von Phraseologismen. Da sich die Schüler häufig nicht der Idiomatizität ihrer Äußerungen bewusst sind, versuchen sie diese wortwörtlich in die Fremdsprache zu übersetzen und beschreiben damit nicht automatisch das von ihnen beabsichtigte. Daher sollen Phraseologismen wie *jemandem den Kopf waschen* zuerst in der Muttersprache bewusst gemacht werden und dann in der Fremdsprache.¹¹³⁶

Wie zu sehen ist, versetzt die genaue Kenntnis des Vokabulars einer Sprache den Fremdsprachenlerner leider noch nicht in die Lage, dieses auch korrekt zu verwenden. Lerner gehen zunächst davon aus, sie könnten einzelne Lexeme beliebig kombinieren, was jedoch nicht der Fall ist. Ein Schüler ist sich oft unsicher darüber, ob er ein bestimmtes Lexem mit einem anderen kombinieren darf oder ob es dabei Restriktionen gibt. Des Weiteren muss er sich fragen, ob es sich bei einer Wortverbindung um eine übliche Kollokation oder eher um eine unübliche beziehungsweise sogar falsche Kombination handelt. Für die Arbeit mit Kollokationen rät Doyé zur Verwendung von Stilwörterbüchern, bei denen zusätzlich zu deutschen und englischen Erklärungen die bevorzugten Kollokationen jedes Lexems aufgeführt sind.¹¹³⁷ Anstatt eines solchen Stilwörterbuchs ist aber auch die Arbeit mit Korpora denkbar, wobei die jeweiligen Wortverbindungen dort zusätzlich auf deren Häufigkeit hin untersucht werden können. Auch die Wörterbücher REDES und PRÁCTICO sowie das Online-Wörterbuch DiCE könnten für die Arbeit mit Kollokationen genutzt werden, doch Erstere sind zum einen keine richtigen Kollokationswörterbücher, zum anderen sind sie relativ teuer in der Anschaffung und werden daher für die Arbeit im Unterricht nicht eigens für Schulen gekauft werden. Das DiCE wäre eine passende und kostenfreie Möglichkeit für die

¹¹³⁵ Doyé, Peter (1971): *Systematische Wortschatzvermittlung im Englischunterricht*. Hannover: Schroedel, S. 88.

¹¹³⁶ Vgl. Hausmann (1984): op. cit., S. 405.

¹¹³⁷ Vgl. Doyé (1971): op. cit., S. 88-91.

Kollokationsarbeit, doch es ist bislang, wie erwähnt, auf die Substantive im Bereich der Gefühle beschränkt. Das HARenEs ist ebenfalls kostenlos und online verfügbar, doch hier kann nur nach der Basis und nicht nach dem Kollokator gesucht werden.

Besonders auf dem Gebiet der Kollokationen wäre die Arbeit mit Korpora demnach lohnenswert. Die Lehrkraft könnte ohne Anschaffung neuer Wörterbücher eine beispielhafte Unterrichtseinheit zu Kollokationen einplanen und den Schülern zudem die Nutzung der Korpora in diesem Bereich erklären. Die Schüler werden im Umgang mit Korpora umso versierter je mehr Unterrichtseinheiten die Lehrkraft anhand derselben durchführt. Daher wäre die regelmäßige Suche nach Kollokationen, beispielsweise am Ende einer Lektion, wünschenswert. Wie die Korpora der spanischen Sprache für die Ermittlung von Kollokationen genutzt werden können, wird im Folgenden beschrieben werden.

4.1.1.2. Kollokationen in Korpora

Wie in Kapitel 1.4.8. zu sehen war, ist bei der Ermittlung von Kollokationen nicht die bloße Häufigkeit ausschlaggebend. Aus diesem Grund spielt der Vergleich der Frequenz der beiden Einheiten an sich mit der Häufigkeit des gemeinsamen Vorkommens als Kollokation eine wichtige Rolle, wobei dieses Verhältnis, wie zu sehen war, anhand von drei verschiedenen Verfahren ermittelt werden kann: MI, *T-Score* und *LL-Simple* (siehe Kapitel 1.4.8.).

Im CORPES der *Real Academia Española* sind eben diese drei Verfahren für die Ermittlung der dort als *Coapariciones* bezeichneten Kombinationen verfügbar. Zusätzlich zum Suchbegriff kann die Wortart angegeben werden, um Homographen zu unterscheiden, und es können die geographische Herkunft sowie das Thema ausgewählt werden. Die Ergebnisse können ihrerseits nach Wortart, Häufigkeit oder einem der drei statistischen Verfahren sortiert werden.¹¹³⁸ Im CORPES wird beispielsweise bei der Anordnung der *Coapariciones* des Substantivs *error* nach dem Wert der MI das Adjektiv *garrafal* an zweiter Stelle insgesamt und als erstes Adjektiv überhaupt angegeben. Man kann in der Liste der *Coapariciones* auch gleich erkennen, dass zwar das auf Platz 34 erscheinende *grave* das am häufigsten mit *error* zusammen vorkommende Adjektiv ist,

¹¹³⁸ Vgl. Real Academia Española: *CORPES. Manual de uso*. Op. cit. (20.10.2017), S. 16.

die höchsten Werte bei MI, *T-Score* und *LL-Simple* allerdings für *garrafal* angegeben werden.

	Clase	Freq	MI	LL SIMPLE	T-SCORE
craso	sustantivo	63	13,19	479,02	7,93
garrafal	adjetivo	97	12,81	698,81	9,84
craso	adjetivo	89	12,57	622,07	9,43
tipeo	sustantivo	10	12,3	67,61	3,16
cometer	verbo	4926	11,53	30.973,37	70,17
ortográfico	adjetivo	66	10,94	382,91	8,12
enmendar	verbo	148	10,74	839,38	12,16
imperdonable	adjetivo	85	10,34	459,85	9,21
muestral	adjetivo	32	10,22	170,64	5,65
subsanan	verbo	77	10,08	404,23	8,77
tipográfico	adjetivo	21	9,92	108,10	4,58
acierto	sustantivo	270	9,81	1.371,30	16,43
corregir	verbo	577	9,58	2.853,23	24,02
margen	sustantivo	572	9,56	2.821,39	23,91
caro	adverbio	43	9,41	207,73	6,55
ortografía	sustantivo	59	9,37	283,38	7,68
incurrir	verbo	167	9,32	797,26	12,92
mayúsculo	adjetivo	20	9,24	94,41	4,47
arbitral	adjetivo	52	9,18	243,58	7,21
errata	sustantivo	12	9,06	55,37	3,46

1 - 20 de 1.736 página: 1 2 3 4 5 6 7 8 9 ... 87

Abbildung 80: CORPES – *Coapariciones* des Substantivs *error*, Ergebnisse 1-20.¹¹³⁹

	Clase	Freq	MI	LL SIMPLE	T-SCORE
gramatical	adjetivo	35	9,04	161,04	5,91
clamoroso	adjetivo	11	9,03	50,53	3,31
omisión	sustantivo	98	8,97	446,70	9,89
medicación	sustantivo	89	8,9	401,79	9,43
rectificar	verbo	80	8,9	361,01	8,94
0.05	sustantivo	12	8,89	54,05	3,46
estándar	adjetivo	58	8,87	260,83	7,61
zaga	sustantivo	31	8,86	139,12	5,56
defensivo	adjetivo	157	8,85	703,89	12,52
desacierto	sustantivo	15	8,71	65,91	3,87
equivocación	sustantivo	48	8,67	209,92	6,92
inconsistencia	sustantivo	22	8,64	95,83	4,69
rectificación	sustantivo	24	8,5	102,45	4,89
grave	adjetivo	865	8,43	3.666,34	29,34
principiante	sustantivo	21	8,33	87,51	4,58
corrección	sustantivo	88	8,3	365,37	9,38
birdie	sustantivo	11	8,3	45,64	3,31
reincidir	verbo	10	8,3	41,48	3,16
involuntario	adjetivo	44	8,26	181,49	6,63
irreparable	adjetivo	23	8,23	94,44	4,79

21 - 40 de 1736 página: 1 2 3 4 5 6 7 8 9 ... 87

Abbildung 81: CORPES – *Coapariciones* des Substantivs *error*, Ergebnisse 21-40.¹¹⁴⁰

Grundsätzlich ist die bloße Frequenz bei Kollokationen zwar keine ausschlaggebende Größe, doch auf dem Gebiet der Korpuslinguistik spielt sie durchaus

¹¹³⁹ Real Academia Española (ed.): *CORPES. Coapariciones*. Online Publikation: <http://web.frl.es/CORPES/org/publico/pages/colocacion/coaparicion.view> (31.08.2018).

¹¹⁴⁰ Real Academia Española (ed.): *CORPES. Coapariciones*. Op. cit. (31.08.2018).

eine Rolle. Zudem ist es für den Fremdsprachenunterricht von Bedeutung, ob bestimmte Kombinationen häufig in Äußerungen von Muttersprachlern vorkommen oder nicht, da es insbesondere die frequenten Verbindungen sind, die im Unterricht vermittelt werden sollten. Aus diesem Grund sollten im Einzelfall die Werte für MI, *T-Score* und *LL-Simple* mit der bloßen Häufigkeit gegeneinander abgewogen werden.

Auch in zwei anderen Korpora der *Real Academia Española* kann nach *Coapariciones* gesucht werden. Hierbei handelt es sich um die annotierte Version des CREA (siehe Kapitel 2.1.1.1.) und das CDH (siehe Kapitel 2.2.1.2.). Bei beiden Korpora können die Ergebnisse nach der Wortart, der Frequenz oder den Werten für MI, *T-Score* und *LL-Simple* geordnet werden.

Im CdE (siehe Kapitel 2.1.1.3., 2.2.1.3. und 2.6.1.) wird zusätzlich zur bloßen Frequenz der Wert der MI angegeben. Der Vorteil des CdE ist dabei die übersichtliche Ergebnisanzeige sowie die Möglichkeit der Darstellung als Statistik. Bei einer Suche nach dem gemeinsamen Auftreten zweier Lexeme, wie beispielsweise nach einem beliebigen Substantiv mit dem Adjektiv *bella*, werden bei gleichzeitiger Auswahl von *Mostrar* im Bereich *Secciones* und von *Relevancia* im Feld *Ordenar* weitere Einzelheiten zur Verwendung des Adjektivs sichtbar:

The screenshot shows the search interface of the CdE:Orig (Version 2008). The search criteria are: PALABRA(S) 'bella', COLOCADOS '[NN*]', CAT GRAM 'nombre.TODOS'. The results are displayed in two columns, labeled 1 and 2, showing a list of sections (s15 to s20) under the heading '-IGNORAR-'. The interface also includes options for ordering (ORDENAR) and limiting (MÍNIMO) the results.

Abbildung 82: CdE:Orig (Version 2008) – Suche nach *bella* mit einem beliebigen Substantiv.¹¹⁴¹

So ist die Übersicht, die sich nach Durchführung dieser Suche im rechten oberen

¹¹⁴¹ Davies, Mark: *Corpus del Español – Género/Histórico (Version 2008)*. Online Publikation: <https://www.corpusdelespanol.org/hist-gen/2008/x.asp> (09.11.2017).

Feld zeigt, um drei Abschnitte erweitert im Vergleich zur gewöhnlichen Ansicht:

CONTEXTO	TODOS	s13	s14	s15	s16	s17	s18	s19	s20	ACAD	PER	FIC	ORAL	SEC1	TODOS	%	IM
DURMIENTE	17							1	16	5	7	3	1	17	48	35.42	11.71
DESPOSADA	5							5						5	78	6.41	9.24
MOLTO	6							6						6	191	3.14	8.21
SINFONÍA	4							2	2		1	1		4	441	0.91	6.42
ESTATUA	7							5	2	1		1		7	868	0.81	6.25
FLOR	11							10	1				1	11	2001	0.55	5.70
IMAGEN	20							19	1		1			20	3857	0.52	5.61
FAZ	5							5						5	1003	0.50	5.56
PERSPECTIVA	5							5						5	1017	0.49	5.54
ÍNDOLE	4							4						4	883	0.45	5.42
CRIATURA	6							6						6	1328	0.45	5.41
DAMA	11							9	2	1	1			11	2526	0.44	5.36
ESPOSA	17							16	1			1		17	4138	0.41	5.28
MUCHACHA	7							1	6		1	5		7	1884	0.37	5.13
PRINCESA	5							5						5	1539	0.32	4.94

Abbildung 83: CdE:Orig (Version 2008) – Ergebnis der Suche nach Substantiven, die zusammen mit *bella* auftreten.¹¹⁴²

Der erste Abschnitt rechts nach der Auflistung der Jahrhunderte und Genres (*TODOS*) gibt die Gesamtanzahl des mit *bella* auftretenden Lexems im Korpus an. Im zweiten Abschnitt (%) wird mittels einer Prozentzahl gezeigt, wie oft dieses Lexem in den angegebenen Jahrhunderten oder Genres mit *bella* im Vergleich zur gesamten Häufigkeit vorkommt. Im dritten Abschnitt (*IM*) wird der Wert der *Información Mutua*, also der MI, angegeben. Soweit der Wert größer oder gleich 3.0 ist, besteht eine semantische Verbindung zwischen den Komponenten. Die Lexeme treten somit aufgrund dieser Verbindung und nicht zufällig häufig zusammen auf.

Außerdem können im CdE Kollokationen verschiedener Lexeme miteinander verglichen werden, so zum Beispiel die mit *mujeres* und *hombres* auftretenden Adjektive.¹¹⁴³

¹¹⁴² Davies: *Corpus del Español – Género/Histórico (Version 2008)*. Op. cit. (09.11.2017).

¹¹⁴³ Vgl. Davies, Mark: *Corpus del Español – Five Minute Tour*. Online Publikation: <http://www.corpusdelespanol.org/> (20.05.2011).

4. Nutzung von Korpora in der fremdsprachlichen Unterrichtspraxis

The screenshot shows the 'COMPARAR' (Compare) tab selected. Under 'BUSCAR' (Search), 'PALABRA(S)' (Word(s)) are 'mujer' and 'hombre'. 'COLOCADOS' (Positions) are '[J*]' and '1'. 'CAT GRAM' (Category Grammar) is 'adj.TODOS'. Below are buttons for 'AL AZAR' (Random), 'BUSCAR' (Search), and 'BORRAR' (Delete). The 'SECCIONES' (Sections) section shows two columns, '1' and '2', each with a list of adjectives: s20, s19, s18, s17, s16, s15. The 'ORDENAR Y LIMITAR' (Order and Limit) section shows 'ORDENAR' (Order) set to 'RELEVANCIA' (Relevance) and 'PALABRA 1 : 2'. 'MÍNIMO' (Minimum) is set to 'FRECUENCIA' (Frequency) with a value of 4. A 'VER OPCIONES' (View Options) button is at the bottom.

Abbildung 84: CdE:Orig (Version 2008) – Vergleich der Adjektive im Umfeld von *mujer* und *hombre*.¹¹⁴⁴

PALABRA 1 (P1): MUJER (0.61)						PALABRA 2 (P2): HOMBRE (1.65)					
	PALABRA	P1	P2	FREC1/FREC2	SCORE		PALABRA	P2	P1	FREC2/FREC1	SCORE
1	HERMOSA	112	0	224.0	369.3	1	HONRADO	250	0	500.0	303.2
2	BUENA	127	1	127.0	209.4	2	BUEN	220	0	440.0	266.9
3	ÚNICA	49	0	98.0	161.6	3	SOLO	153	0	306.0	185.6
4	SOLA	48	0	96.0	158.3	4	MODERNO	86	0	172.0	104.3
5	HONRADA	47	0	94.0	155.0	5	PÚBLICO	67	0	134.0	81.3
6	AMADA	45	0	90.0	148.4	6	NUEVO	57	0	114.0	69.1
7	PROPIA	36	0	72.0	118.7	7	ÚNICO	52	0	104.0	63.1
8	MALA	35	0	70.0	115.4	8	ALTO	46	0	92.0	55.8
9	BELLA	34	0	68.0	112.1	9	POLÍTICO	46	0	92.0	55.8
10	DESNUDA	33	0	66.0	108.8	10	CULTO	38	0	76.0	46.1
11	ESPAÑOLA	25	0	50.0	82.4	11	GORDO	29	0	58.0	35.2
12	LEGÍTIMA	25	0	50.0	82.4	12	RUDO	28	0	56.0	34.0
13	MISTERIOSA	25	0	50.0	82.4	13	MALO	26	0	52.0	31.5
14	DIVINA	18	0	36.0	59.4	14	PRÁCTICO	26	0	52.0	31.5
15	GORDA	17	0	34.0	56.1	15	SERIO	26	0	52.0	31.5
16	DESVENTURADA	16	0	32.0	52.8	16	VIRTUOSO	26	0	52.0	31.5
17	EMBARAZADA	16	0	32.0	52.8	17	DIGNO	25	0	50.0	30.3
18	BLANCA	16	0	32.0	52.8	18	VERDADERO	25	0	50.0	30.3

Abbildung 85: CdE:Orig (Version 2008) – Ergebnis des Vergleichs der Adjektive im Umfeld von *mujer* und *hombre*.¹¹⁴⁵

Auch ein Vergleich der Kollokationen mit ein und demselben Lexem als Basis in verschiedenen zeitlichen Abschnitten oder innerhalb unterschiedlicher Register ist möglich.¹¹⁴⁶

4.1.1.3. Erstellung von Unterrichtsmaterial zum Thema Kollokationen

Die eben aufgeführten Funktionen können nun genutzt werden, um Unterrichtsmaterial zum Thema Kollokationen zu erstellen. Ein paar dieser Möglichkeiten sind im Folgenden dargestellt und erklärt.

¹¹⁴⁴ Davies: *Corpus del Español – Género/Histórico (Version 2008)*. Op. cit. (09.11.2017).

¹¹⁴⁵ Davies: *Corpus del Español – Género/Histórico (Version 2008)*. Op. cit. (09.11.2017).

¹¹⁴⁶ Vgl. Davies: *Corpus del Español – Five Minute Tour*. Op. cit. (20.05.2011).

4.1.1.3.1. Erweiterung der Ausdrucksfähigkeit anhand von Kollokationen

Wie in Kapitel 4.1.1.2. zu sehen war, können im Korpus CORPES (siehe Kapitel 2.1.1.2.) Kollokationen anhand verschiedener statistischer Verfahren ermittelt werden. Mittels des Arbeitsblatts *Colocaciones* (AB 1a in Anhang 8) sollen die Schüler beispielsweise anhand realer Sätze aus dem CORPES selbstständig Kollokationen erarbeiten.¹¹⁴⁷ Bei den hier aufgeführten Kollokationen handelt es sich um solche Kombinationen, die meist nicht oder nur unter dem Kollokator in Wörterbüchern zu finden sind. Kollokationen, welche dort unter der Basis geführt werden, können im Unterricht auch mithilfe der gewöhnlichen Wörterbucharbeit ermittelt werden, die hier angeführten jedoch nicht, da sie nur durch systematische Suche im Nachschlagewerk zu finden sind.

Um die Kollokatoren zu finden, geben die Schüler im Reiter *Coapariciones* das Lemma der Basis ein (zum Beispiel *problema*) und wählen die entsprechende Wortart aus (*sustantivo*). Gegebenenfalls muss die Lehrkraft die Schüler darauf hinweisen, dass die Lexeme in der Grundform, also im Singular beziehungsweise im Infinitiv, einzugeben sind. Durch einen Klick auf die Option *Coapariciones* erhalten sie im Folgenden eine Liste derjenigen Lexeme, die bevorzugt mit der Basis kombiniert werden, wobei für die Reihenfolge der Aufzählung der Wert der MI ausschlaggebend ist (siehe Kapitel 1.4.8.).

¹¹⁴⁷ Die für die vorliegende Arbeit erstellten Arbeitsblätter sind so gestaltet wie die ersten Kapitel in Lehrwerken für den Anfangsunterricht. Die Titel und allgemeine, als bekannt vorauszusetzende Angaben werden aufgrund der im Fremdsprachenunterricht anzustrebenden Einsprachigkeit auf Spanisch verfasst, doch die auf das jeweilige Korpus bezogene Aufgabenstellung ist auf Deutsch formuliert, um das Verständnis der Lerner zu sichern. Für fortgeschrittene Lerner, die bereits Erfahrung im Umgang mit Korpora gesammelt haben, könnten die Arbeitsaufträge durchaus auf Spanisch formuliert werden. Obwohl in der vorliegenden Arbeit Wert darauf gelegt wurde, anstatt des umgangssprachlichen Begriffs ‚Wort‘ den genaueren und wissenschaftlicheren Terminus ‚Lexem‘ zu verwenden, wird in den Arbeitsblättern dennoch auf Ersteren zurückgegriffen, da sie für Schüler formuliert wurden und diesen der Begriff ‚Lexem‘ in der Regel nicht bekannt ist.

4. Nutzung von Korpora in der fremdsprachlichen Unterrichtspraxis

Concordancias Coapariciones Configuración Ayuda Modo de cita Sugerencias				
Lema:	problema	Clase de palabra:	sustantivo	
Origen	(Todos) América España Filipinas			
Coapariciones Nueva consulta				
	Clase	Freq	MI	LL SIMPLE
solucionar	verbo	3254	9,83	17.143,12
acuciante	adjetivo	133	9,32	650,06
irresoluble	adjetivo	39	9,09	184,10
resolver	verbo	7080	8,98	33.072,04
aquejar	verbo	303	8,83	1.379,27
broncopulmonar	adjetivo	11	8,72	49,18
solventar	verbo	279	8,63	1.232,56

Abbildung 86: CORPES – Suche nach Kollokationen mit dem Substantiv *problema*.¹¹⁴⁸

Aus dieser Liste wählen die Schüler einen der Wortart nach passenden Kollokator aus, wobei die jeweilige Wortart in der Aufgabe angegeben ist. Bei leistungsstarken Schülern könnte diese Angabe weggelassen werden. Für die Auswahl des Kollokators kann die Frequenz als zusätzliches Kriterium herangezogen werden, doch auf dem Gebiet der Kollokationen sind, wie bereits mehrfach erwähnt wurde, vor allem die Angaben der statistischen Verfahren MI, *LL-Simple* oder *T-Score* zu berücksichtigen. Im ersten Beispielsatz sollte somit das Verb *solucionar* ausgewählt und in der richtigen Form (in diesem Fall als Infinitiv) eingesetzt werden. Auch das Verb *resolver* wäre hier möglich, da dafür eine sehr hohe Frequenz und ein nur geringfügig niedrigerer MI-Wert angegeben wird. Zur selbstständigen Korrektur sowie zur Auflockerung der Aufgabe können die Lösungen zusätzlich in ein Kreuzworträtsel eingetragen werden (AB 1b in Anhang 9).¹¹⁴⁹ Das Kreuzworträtsel für diese Aufgabe wurde mit dem im Internet kostenlos nutzbaren Programm *XWords* erstellt.¹¹⁵⁰ Der Vorteil bei der Nutzung des Kreuzworträtsels ist die Möglichkeit der selbstständigen Kontrolle durch die Schüler sowie der zusätzliche spielerische Effekt. Der Nachteil ist allerdings, dass es folglich nur eine korrekte Lösung gibt und somit das im genannten Beispielsatz ebenso geeignete *resolver* nicht von den Schülern verwendet werden kann. Hier kann je nach Klasse oder Kurs abgewägt werden,

¹¹⁴⁸ Real Academia Española (ed.): *CORPES. Coapariciones*. Op. cit. (31.08.2018).

¹¹⁴⁹ Idee der Kombination mit einem Kreuzworträtsel aus: Montoro del Arco (2010): op. cit., S. 253-269.

¹¹⁵⁰ Vgl. Ebert, Lars (2012-2017): *XWords*. Online Publikation: <https://www.xwords-generator.de/de> (26.10.2017).

ob diese spielerische, aber doch sehr lenkende Komponente eingebaut wird oder ob die Aufgabe freier gestaltet werden soll.

Die Beispielsätze in Arbeitsblatt 1 sind allesamt dem CORPES entnommen, um den Schülern einen realen sprachlichen Kontext für die Kollokationen aufzuzeigen, wobei kleinere Änderungen eingefügt und die Sätze teilweise gekürzt wurden. Dies bedeutet wiederum, dass das sprachliche Niveau der Übung sehr hoch angesetzt ist und diese daher frühestens ab Ende des dritten Lernjahrs zum Einsatz kommen sollte. Generell setzt das Lernen und vor allem die Verwendung von Kollokationen bereits eine gewisse sprachliche Kompetenz voraus. Übungen wie auf diesem Arbeitsblatt sind daher insbesondere zur Erweiterung der sprachlichen Fähigkeiten im Hinblick auf die Textproduktion geeignet. Kollokationen wie *esgrimir argumentos*, *adoptar medidas*, *implementar estrategias* oder auch *negarse rotundamente* und *de suma importancia* sind schließlich vor allem für das Verfassen von Aufsätzen nützlich, da sie im rezeptiven Bereich keine großen Probleme bereiten.

Aufgrund des hohen sprachlichen Niveaus, welches durch den Umgang mit einem fremdsprachlichen Korpus wie dem CORPES, das aus realen Texten aus dem spanisch- und lateinamerikanischsprachigen Ausland besteht, angesetzt wird, sind den Schülern jedoch nicht alle auf dem Arbeitsblatt verwendeten Ausdrücke und zudem auch einige im CORPES aufgelisteten Kollokatoren nicht bekannt. Aus diesem Grund sollten die Schüler mit einem Online-Wörterbuch arbeiten, beispielsweise von PONS (<https://de.pons.com>). PONS-Wörterbücher sind an Schulen weitverbreitet und den Schülern daher meist bereits bekannt.

Der sprachliche Kontext des Arbeitsblatts *Colocaciones* ist durch die Verwendung realer Beispiele aus dem Korpus durchaus fordernd und verlangt profunde Spanischkenntnisse von den Schülern. Sofern eine Lehrkraft sich zusammen mit Schülern mit Kollokationen beschäftigen möchte, jedoch die Komplexität der Beispielsätze scheut, kann sie Arbeitsblatt 2a verwenden (Anhang 10). Hier sollen die Schüler die Kollokationen im CORPES suchen, der sprachliche Kontext der Beispielsätze ist jedoch nur an die Konkordanzen angelehnt und speziell für Schüler mit geringeren Sprachkenntnissen einfach gehalten. Zudem sollen die Schüler die Bedeutung der jeweiligen Kollokation anhand eines Online-Wörterbuchs ermitteln und angeben. Dieses Arbeitsblatt ist schon ab Ende des zweiten oder zu Beginn des dritten Lernjahrs geeignet.

Egal ob das Arbeitsblatt 1 nun mit oder ohne Kreuzworträtsel oder auch das Arbeitsblatt 2a verwendet wird, so handelt es sich doch um eine sehr gelenkte Übung. Derart gelenkte Vorgehensweisen haben durchaus ihre Berechtigung im Fremdsprachenunterricht, sie bieten dem Lerner einen passenden sprachlichen Kontext für die neuen Lexeme und vermitteln ihm Sicherheit, da genau vorgegeben wird, was zu tun ist. Dennoch sollen die Aufgabenformen generell zunehmend freier gestaltet werden, insbesondere für fortgeschrittene Lerner. Eine freiere Aufgabenform zur Beschäftigung mit Kollokationen bietet Arbeitsblatt 2b (Anhang 11), welches ebenso ab Ende des zweiten Lernjahrs genutzt werden kann. Hier ist keinerlei sprachlicher Kontext gegeben, sondern die Schüler sollen selbst mithilfe der Konkordanzen des CORPES und eines Online-Wörterbuchs Beispielsätze erarbeiten. Die Lösung zu diesem Arbeitsblatt kann daher nur ein Lösungsvorschlag sein, da die Schüler frei im Korpus suchen dürfen und auch nicht den Kollokator mit dem größten MI-Wert auswählen müssen. Sie können ebenso den dritten oder vierten Treffer der jeweiligen Wortart angeben und diese Kollokation für ihren Beispielsatz verwenden. Die freie Auswahl und die verschiedenen Lösungen erschweren dem Lehrer zwar die Verbesserung des Arbeitsblatts, sie motivieren die Schüler jedoch zusätzlich. Zudem bieten mehrere Kollokationen einen weiteren Vorteil: Je mehr Kollokationen und sprachliche Kontexte den Schülern vorliegen, desto besser können sie das Vokabular verankern. Unabhängig davon, für welches Arbeitsblatt sich die Lehrkraft entscheidet, ob für eine freiere oder gelenktere Vorgehensweise, grundsätzlich kann gesagt werden, dass es den Schülern leicht fällt, Kollokationen zu erkennen und diese in das jeweilige Arbeitsblatt einzusetzen.

Es muss jedoch hinzugefügt werden, dass diejenigen Kollokationen, welche die Schüler anhand des CORPES suchen sollen, von der Lehrkraft zuvor getestet werden sollten, da die Auflistung der Lexeme nach dem Wert der MI in der Funktion *Coapariciones* in allen Korpora der *Real Academia Española* bisweilen eigenartige Ergebnisse liefern kann. Sucht man beispielsweise nach dem Begriff *dictadura*, so finden sich in der Auflistung nach dem Wert der MI unter den ersten Ergebnissen Jahreszahlen, einige Namen wie Primo de Rivera, Marcos Pérez Jiménez oder Augusto Pinochet sowie die Adjektive *somocista*, *franquista* und *castrista*. Im CREA (siehe Kapitel 2.1.1.1.) und im CDH (siehe Kapitel 2.2.1.2.) fallen die Ergebnisse ähnlich aus, abgesehen von den Jahreszahlen. Die gefundenen Ergebnisse sind ohne Zweifel mit dem Begriff *dictadura*

verbunden, doch sie werfen bei den Schülern mehr Fragen auf als sie zur Wortschatzerweiterung beitragen. Durch die Auflistung der Lexeme nach dem Wert der MI werden vor allem diejenigen Lexeme weit oben angesiedelt, die eine besonders enge Beziehung zum Suchbegriff verbindet. Im Hinblick auf die genannten Jahreszahlen und Namen ist die Beziehung wohl deshalb so eng, weil diese außerhalb des Kontexts *dictadura* nicht im Korpus vorkommen.

Im Fall der Arbeitsblätter 1a bis 2b handelt es sich um sorgfältig ausgewählte Kollokationen, welche ohne Probleme anhand des CORPES ermittelt werden können. Wenn Lehrkräfte allerdings andere Kollokationen anhand des CORPES zusammen mit den Schülern ermitteln möchten, so sollten sie die Suchanfragen zuvor selbst testen.

Aus diesem Grund ist für die selbstständige Beschäftigung der Schüler mit dem CORPES oder auch dem CREA die Auflistung nach den Werten der Verfahren *T-Score* oder *LL-Simple* empfehlenswerter. Wie bereits in Kapitel 1.4.8. erklärt wurde, beziehen diese beiden Werte auch die Anzahl der gefundenen Belege mit in ihre Berechnung ein und berücksichtigen somit vermehrt frequente Kombinationen.

Wenn die Lehrkraft die Schüler beispielsweise im CORPES nach Kombinationen von *dictadura* oder *gobierno* mit einem Adjektiv suchen lassen möchte, so sollte sie den Schülern dazu raten, die Ergebnisse nach dem Wert *T-Score* oder nach *LL-Simple* anordnen zu lassen. Wird allerdings ausschließlich nach diesen Werten sortiert, so führen Funktionswörter die Liste der *Coapariciones* an. Aus diesem Grund sollte beispielsweise eine Kombination aus *T-Score* und *Clase de Palabra* gewählt werden. Klickt der Nutzer zunächst auf *T-Score*, sodass diejenigen Lexeme mit dem höchsten Wert als erstes angezeigt werden, und danach auf *Clase de Palabra*, so werden die Ergebnisse zudem nach der Wortart aufgelistet. Folglich stehen die Adjektive mit dem größten *T-Score*-Wert ganz oben. Sucht man nach Substantiven zu einem Verb, müssen entsprechend die weiteren Ergebnisseiten durchgesehen werden. Bei Lexemen, welche sehr häufig vorkommen, wie beispielsweise bei *tener*, gibt es jedoch über 1000 Trefferseiten, weshalb der Nutzer erst einmal einige Seiten durchblättern muss, um zu den Substantiven zu gelangen.

Um die Unterschiede der einzelnen statistischen Verfahren zur Ermittlung von Kollokationen zu veranschaulichen, wird im Folgenden exemplarisch nach dem Verb *tener* gesucht:

4. Nutzung von Korpora in der fremdsprachlichen Unterrichtspraxis

	Clase	Freq	MI ▼	LL SIMPLE	T-SCORE
maltraer	verbo	22	7,53	90,70	4,69
derecho	adverbio	108	7,47	438,99	10,39
parangón	sustantivo	125	7,18	474,45	11,18
seudobulbo	sustantivo	15	7,17	56,78	3,87
en común	adverbio	2193	7,1	8.153,24	46,50
cabida	sustantivo	917	7,01	3.331,98	30,05
empacho	sustantivo	162	6,9	575,82	12,64
rolo	adjetivo	10	6,71	33,97	3,16
repercusión	sustantivo	1636	6,67	5.509,57	40,07
tirria	sustantivo	34	6,65	113,98	5,83
agalla	sustantivo	147	6,55	482,07	12,04
equipazo	sustantivo	24	6,53	78,37	4,89
encontrón	sustantivo	10	6,53	32,68	3,16
ojeriza	sustantivo	34	6,52	110,66	5,83
baraka	sustantivo	11	6,52	35,83	3,31
correlato	sustantivo	160	6,5	519,46	12,57
presentimiento	sustantivo	381	6,45	1.221,16	19,31
viso	sustantivo	202	6,45	648,15	14,07
chance	sustantivo	668	6,37	2.101,75	25,53
gana	sustantivo	5716	6,35	17.906,11	74,69

1 - 20 de 21.488 página: 1 2 3 4 5 6 7 8 9 ... 1075

Abbildung 87: CORPES – Auflistung der ersten 20 *Coapariciones* des Verbs *tener* angeordnet nach dem Wert der MI.¹¹⁵¹

So findet sich bei Anordnung der *Coapariciones* des Verbs *tener* rein nach dem Wert der MI an erster Stelle das Verb *maltraer* mit 22 gemeinsamen Vorkommen. An zweiter Stelle steht *derecho* mit 108 Treffern. An dritter und vierter Stelle stehen die Substantive *parangón* und *seudobulbo*. Wenn man sich allerdings die Konkordanzen zu diesen beiden Kollokationen anzeigen lassen möchte, kann das Korpus keine Belege finden. An fünfter Stelle steht schließlich das Adverb *en común* mit 2193 gemeinsamen Vorkommen, welches man durchaus weit oben in der Auflistung vermutet hätte. Die Ergebnisse sind daher teilweise unverständlich und können nur dadurch erklärt werden, dass diese niedrig frequenten Lexeme ausschließlich im Kontext von *tener* im Korpus vorkommen, weshalb sie einen sehr hohen Wert im Verfahren der MI erzielen. Teilweise scheint die Option *Coapariciones* im CORPES allerdings noch nicht reibungslos zu funktionieren, wenn bei manchen Treffern keine Belege angezeigt werden können. In der vorherigen Version 0.83 umfasste das CORPES zwar weniger Lexeme und verfügte noch nicht über alle derzeitigen Suchoptionen, das eben genannte Problem zur fehlerhaften Anzeige von Konkordanzen gab es in jener Version allerdings noch nicht. Es bleibt zu hoffen, dass diese fehlerhaften Auflistungen in einer der folgenden Versionen verbessert werden.

¹¹⁵¹ Real Academia Española (ed.): *CORPES. Coapariciones*. Op. cit. (31.08.2018).

Werden die *Coapariciones* jedoch nach dem Wert der Verfahren *LL-Simple* oder *T-Score* angeordnet, erscheinen fast ausschließlich Funktionswörter wie *el, que, de, uno* oder *y* unter den ersten Treffern. Kombiniert man nun die Auflistung nach einem der statistischen Verfahren mit derjenigen nach der Wortart, so werden relevante Kombinationen speziell nach der jeweiligen Wortart aufgelistet. Die Substantive, welche nach dem Wert der MI hauptsächlich zusammen mit *tener* vorkommen, überzeugen zwar auch hier nicht (*parangón, seudobulbo, cabida, empacho, repercusión, tirria, agalla*), doch die Treffer, welche nach den Verfahren *T-Score* oder *LL-Simple* in Kombination mit der Wortart aufgelistet werden, entsprechen denjenigen, welche man auch als relevante Kombinationen einordnen würde: *cuenta, año, razón, lugar, hijo, problema, miedo, tiempo, oportunidad*.

All diese Suchoptionen im Bereich *Coapariciones* bieten auch die annotierte Version des CREA (siehe Kapitel 2.1.1.1.) und das CDH (siehe Kapitel 2.2.1.2.). Für ähnliche Suchanfragen oder auch für die Bearbeitung der Arbeitsblätter könnten diese beiden Korpora somit ebenso gut verwendet werden. Dennoch wurde in dieser Arbeit das CORPES ausgewählt, da es sich hier um das aktuellste Korpus handelt.

4.1.1.3.2. Erstellung einer Mindmap mithilfe von Korpora

Eine relativ einfache und anschauliche Aufgabe, welche sich bereits für Schüler im Anfangsunterricht eignet, ist die Erstellung einer Mindmap mit Kollokationen. Ein Beispiel für das Themengebiet ‚Aussehen‘ im ersten Lernjahr ist die Erstellung einer Mindmap zum Lexem *pelo* anhand des CdE:Orig (Version 2008). Wie bereits in Kapitel 2.1.1.3. angegeben, ist das alte Interface des CdE:Orig benutzerfreundlicher und intuitiver zu bedienen als das neue. Außerdem werden Suchen schneller durchgeführt und es werden keine Wartemeldungen angezeigt. Aus diesem Grund wird für die hier beschriebene Erstellung einer Mindmap die alte Benutzeroberfläche aus dem Jahr 2008 verwendet:

4. Nutzung von Korpora in der fremdsprachlichen Unterrichtspraxis

MOstrar

☒ LISTA ☐ GRÁFICO ☐ PCEC ☐ COMPARAR

BUSCAR

PALABRA(S) [pelo]

COLOCADOS [J*] 4 4

CAT GRAM adj.TODOS

ALAZAR

BUSCAR BORRAR

SECCIONES ☐ MOstrar

1 -IGNORAR- s20 s19 s18 s17 s16 s15

2 -IGNORAR- s20 s19 s18 s17 s16 s15

ORDENAR Y LIMITAR

ORDENAR FRECUENCIA

MÍNIMO FRECUENCIA 4

VER OPCIONES

Abbildung 88: CdE:Orig (Version 2008) – Einstellungen für die Suche nach Kollokationen bestehend aus dem Substantiv *pelo* und einem Adjektiv.¹¹⁵²

Bei der Suche nach [pelo] mit einem Adjektiv als Kollokation ergeben sich folgende Treffer:

		CONTEXTO	TOT
1	<input type="checkbox"/>	NEGRO	96
2	<input type="checkbox"/>	LARGO	75
3	<input type="checkbox"/>	BLANCO	53
4	<input type="checkbox"/>	RUBIO	42
5	<input type="checkbox"/>	SUELTO	26
6	<input type="checkbox"/>	RIZADO	25
7	<input type="checkbox"/>	CASTAÑO	22
8	<input type="checkbox"/>	REVUELTO	22
9	<input type="checkbox"/>	LACIO	20
10	<input type="checkbox"/>	ERIZADO	19
11	<input type="checkbox"/>	LARGOS	19
12	<input type="checkbox"/>	POCO	19
13	<input type="checkbox"/>	ROJO	18

Abbildung 89: CdE:Orig (Version 2008) – Ergebnisse der Suche nach Kollokationen von *pelo* in Verbindung mit einem Adjektiv.¹¹⁵³

Es kann daraufhin eine Mindmap zum Thema *pelo* erstellt werden, bei der die Schüler die Ergebnisse aus dem Korpus selbst auswählen und anordnen:

¹¹⁵² Davies: *Corpus del Español – Género/Histórico (Version 2008)*. Op. cit. (09.11.2017).

¹¹⁵³ Davies: *Corpus del Español – Género/Histórico (Version 2008)*. Op. cit. (09.11.2017).

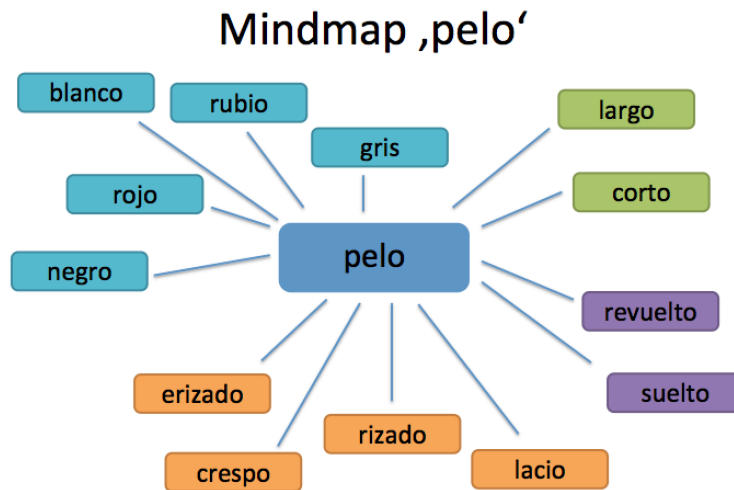


Abbildung 90: Mindmap zum Thema ‚pelo‘.

Wie in Kapitel 4.1.1.3.1. zu sehen war, eignet sich insbesondere das CORPES für die Suche nach Kollokationen. Durch die Anordnung nach dem Wert des statistischen Verfahrens *LL-Simple* und der Wortart können anhand des CORPES auch Mindmaps von Schülern erstellt werden. Auf Arbeitsblatt 3 – *Mapa mental: Aspecto* (siehe Anhang 12), welches ab Mitte des zweiten Lernjahrs genutzt werden kann, findet sich eine genaue Erklärung der Vorgehensweise sowie der Beginn der Mindmap. Im Folgenden steht es den Schülern frei, welche Lexeme sie integrieren und welche Begriffe sie als weitere Basen für die Suche im CORPES nutzen. Die Basis der Kollokation (beispielsweise *pelo*) wird als Lemma eingegeben und die Wortart derselben ausgewählt (*sustantivo*). Wenn sich der Nutzer die Ergebnisse nach dem statistischen Verfahren *LL-Simple* und der Wortart sortiert anzeigen lässt, findet er, wie bereits beschrieben, alle Kollokatoren einer Wortart untereinander aufgereiht und kann diese vergleichen. Anhand dieser Auflistung kann eine Mindmap erstellt werden, wobei nicht nur Adjektive berücksichtigt werden sollen, sondern auch andere Wortarten wie Verben oder Substantive. In der Auflistung der Substantive finden sich zudem weitere Oberbegriffe zum Thema ‚Aussehen‘ (zum Beispiel *ojos*, *piel*, *cara*, *barba*, *oreja* oder *ropa*), welche ihrerseits als Basis dienen können. Mittels eines Klicks auf ein blau geschriebenes Lexem werden Konkordanzen zu dieser Kombination (beispielsweise *pelo largo*) angezeigt, anhand derer die Schüler sehen können, wie die beiden Lexeme genutzt werden und ob sie direkt nebeneinander erscheinen oder in einem weiteren sprachlichen Kontext. Lexeme, deren Bedeutung den Schülern nicht bekannt ist, können in einem Online-Wörterbuch nachgeschlagen werden.

Auch mithilfe des opportunistischen Korpus *Spa_newscrawl_2011* innerhalb der Leipziger Korpussammlung (siehe Kapitel 2.1.3.2.) kann eine Art Übersicht erstellt werden, die dort als Graph bezeichnet wird. Dabei handelt es sich um eine graphische Darstellung von Kookkurrenzen zum Suchbegriff (zum Begriff Kookkurrenz siehe auch Kapitel 4.1.1.). Der Graph zum Lexem *cielo* sieht beispielsweise wie folgt aus:

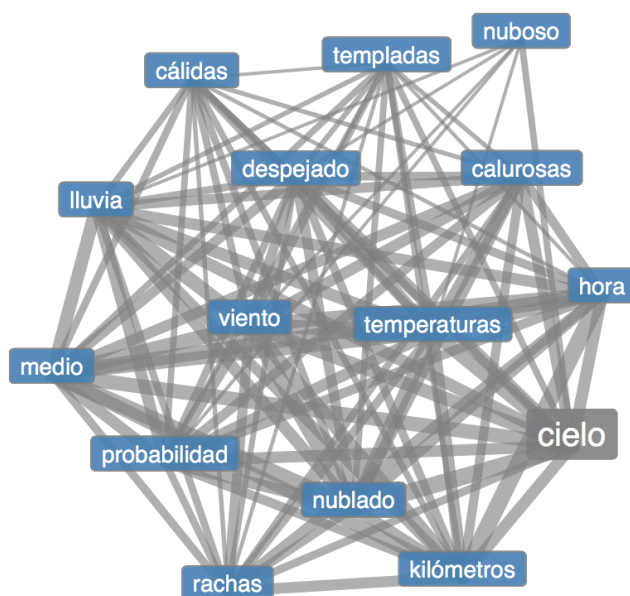


Abbildung 91: *Spa_newscrawl_2011* – Graph zum Substantiv *cielo*.¹¹⁵⁴

Hier handelt es sich zwar um Kookkurrenzen und keine Kollokationen, doch einige der aufgeführten Lexeme bilden mit dem Substantiv *cielo* ebenfalls eine Kollokation. Die Graphen, die das Korpus *Spa_newscrawl_2011* bereitstellt, können somit durchaus zur Erstellung einer Mindmap genutzt werden. So kann der oben abgebildete Graph zum Thema *cielo* zusammen mit anderen Graphen, beispielsweise zu Lexemen wie *sol*, *lluvia* oder *viento*, als Grundlage für eine Mindmap zum Thema ‚Wetter‘ dienen. Zudem sind im jeweiligen Eintrag signifikante linke und rechte Nachbarn des Suchbegriffs, weitere Kookkurrenzen mit Frequenzangaben sowie Beispielsätze aufgelistet. All diese Angaben verhelfen dem Nutzer dazu, sich über den Suchbegriff, insbesondere über dessen sprachlichen Kontext, zu informieren. Jedoch sind die Beispielsätze nicht im typischen KWIC-Format angeordnet (siehe Kapitel 1.4.3.), da der jeweilige Suchbegriff nicht zentral in jedem Satz untereinander aufgelistet wird, was diesen Teil des Eintrags etwas

¹¹⁵⁴ Universität Leipzig - Institut für Informatik - Abteilung ASV (1998-2017): *Wortschatz Universität Leipzig. Spa_newscrawl_2011. Suche nach ‚cielo‘*. Online Publikation: http://corpora.uni-leipzig.de/de/res?corpusId=spa_newscrawl_2011&word=cielo (10.11.2017).

unübersichtlich erscheinen lässt. Die Kookkurrenzen werden allerdings nur aufgelistet beziehungsweise im Graph dargestellt, aber nicht als Beispielsatz in Form einer Konkordanz angegeben. Dadurch sieht der Nutzer zwar auf den ersten Blick, welche Lexeme häufig mit dem Suchbegriff zusammen vorkommen, doch wie sie zusammen genutzt werden, muss für jede Kookkurrenz eigens ermittelt werden.

Wenn der Nutzer über den vom Korpus *Spa_newscrawl_2011* erstellten Graphen hinaus nach weiteren Kookkurrenzen für die Erstellung einer Mindmap oder zum besseren Verständnis des gewöhnlichen Kontexts eines Begriffs suchen möchte, kann er das *NOW Corpus: Spanish* (siehe Kapitel 2.6.3.) zurate ziehen. In diesem neuen Korpus, welches monatlich erweitert werden soll, können auf relativ einfache Art und Weise sogenannte *virtuelle Korpora* für eigene Nachforschungen erstellt werden.

Um beim Thema ‚Wetter‘ zu bleiben, soll im *NOW Corpus: Spanish* nach dem Substantiv *viento* gesucht werden. Hierfür wird in der Suchmaske der Suchbegriff eingegeben, im Bereich *Virtual* die Option ‚*Encontrar textos*‘ ausgewählt und im Bereich *Orden/Límite* nach Relevanz sowie einer Mindesthäufigkeit von fünf Vorkommen pro Text sortiert. Auf diese Weise werden lediglich diejenigen Texte aufgelistet, in denen der Begriff *viento* mindestens fünf Mal vorkommt. Im nächsten Schritt können diejenigen Texte, welche in das virtuelle Korpus integriert werden sollen, ausgewählt und daraufhin gespeichert werden. Schließlich kann sich der Nutzer die *palabras clave* der verschiedenen Wortarten anzeigen lassen. Für den Suchbegriff *cielo* erscheinen beispielsweise folgende Substantive zu Beginn der Liste des erstellten virtuellen Korpus:

VIENTO [25,633 PALABRAS, 100 TEXTOS] (0.0% DE TODO) **NOUN** VERB ADJ ADV N+N ADJ+N [\[TODOS LOS CORPUS\]](#) [GUARDAR LISTA](#)

AYUDA	PALABRA (HACER CLIC)	FREC	# TEXTOS	ESPECÍFICO FREC 15 10 TEXTOS	CORPUS ENTERO	ESPERADO
1	RÁFAGA	88	32	2,547.4	8,086	0.0
2	VIENTO	750	100	1,067.7	164,424	0.7
3	NEVADA	24	11	657.8	8,540	0.0
4	KM/H	134	24	630.9	49,716	0.2
5	RACHA	141	41	369.9	89,233	0.4
6	KILÓMETRO	92	38	173.2	124,357	0.5
7	NARANJA	48	16	139.0	80,853	0.3
8	NIEVE	39	11	100.4	90,959	0.4
9	ALERTA	68	26	93.8	169,756	0.7
10	TEMPERATURA	53	29	77.8	159,362	0.7

Abbildung 92: *NOW Corpus: Spanish* – Liste der mit dem Substantiv *viento* erscheinenden Substantive in einem selbst erstellten virtuellen Korpus mit dem Titel ‚*viento*‘.¹¹⁵⁵

¹¹⁵⁵ Davies: *NOW Corpus: Spanish*. Op. cit. (25.08.2018).

4. Nutzung von Korpora in der fremdsprachlichen Unterrichtspraxis

Für das Substantiv *lluvia* erscheinen folgende Adjektive zu Beginn der Auflistung eines neuen virtuellen Korpus mit dem gleichnamigen Titel:

LLUVIA [25,284 PALABRAS, 100 TEXTOS] (0.0% DE TODO) NOUN VERB **ADJ** ADV N+N ADJ+N [\[TODOS LOS CORPUS\]](#) [GUARDAR LISTA](#)

AYUDA	PALABRA (HACER CLIC)	FREC	# TEXTOS	ESPECÍFICO FREC 15 10 TEXTOS	CORPUS ENTERO	ESPERADO
1	LIGERO	26	16	128.1	48,152	0.2
2	INTENSO	33	18	77.7	100,798	0.4
3	ELÉCTRICO	18	10	35.3	121,039	0.5
4	MÁXIMO	23	11	32.4	168,591	0.7
5	FUERTE	69	37	24.0	682,218	2.9
6	DIFERENTE	16	11	11.7	323,358	1.4
7	BAJO	15	13	10.9	326,964	1.4
8	GRANDE	22	18	9.7	536,254	2.3
9	BUENO	21	12	9.3	536,781	2.3
10	MEJOR	20	10	1.9	2,489,692	10.5

Abbildung 93: *NOW Corpus: Spanish* – Liste der mit dem Substantiv *lluvia* erscheinenden Adjektive in einem selbst erstellten virtuellen Korpus mit dem Titel „*lluvia*“.¹¹⁵⁶

Ein weiterer Vorteil der Suche nach Kookkurrenzen innerhalb des *NOW Corpus: Spanish* ist die Anzeige passender Konkordanzen. Im Gegensatz zum *Spa_newscrawl_2011* kann jede der erscheinenden Kookkurrenzen angeklickt werden, um zu einer Auflistung von Konkordanzen zu gelangen, anhand derer der Nutzer erkennen kann, wie beide Lexeme im Kontext verwendet werden. Die Erstellung virtueller Korpora ist auch im CdE:Orig (siehe Kapitel 2.1.1.3. und 2.2.1.3.) und im CdE:Nuevo (siehe Kapitel 2.6.1.) möglich, welche allerdings nicht über derart viele Texte verfügen.

4.1.1.3.3. Erarbeitung eines Kollokationsfelds mithilfe von Korpora

Mit fortgeschrittenen Lernern kann nicht nur eine Mindmap zum Thema *pelo* erstellt werden, wie in Kapitel 4.1.1.3.2. veranschaulicht, sondern sogar ein Kollokationsfeld im Sinne von Doyés zuvor angegebener Definition (siehe Kapitel 4.1.1.1.).¹¹⁵⁷ Hierbei wird im CORPES im Reiter *Coapariciones* das gesuchte Lemma (*pelo*) eingegeben sowie die Wortart (*sustantivo*) ausgewählt. Im Folgenden erscheinen in der Grundeinstellung, wie bereits bekannt, diejenigen Lexeme, die sich aufgrund ihres gemeinsamen Vorkommens mit dem Suchbegriff durch eine besondere Beziehung auszeichnen, geordnet nach dem Wert der MI:

¹¹⁵⁶ Davies: *NOW Corpus: Spanish*. Op. cit. (25.08.2018).

¹¹⁵⁷ Vgl. Doyé (1971): op. cit., S. 88.

	Clase ▲(4)	Freq	MI ▼	LL SIMPLE ▼(3)	T-SCORE ▼(2)
tomadura	sustantivo	153	13,59	0,00	12,36
a lo garçon	adverbio	14	12,99	102,85	3,74
enrulado	adjetivo	37	12,57	257,97	6,08
indumento	sustantivo	23	12,51	159,18	4,79
al rape	adverbio	41	12,39	279,81	6,40
lacio	adjetivo	280	12,37	1.907,16	16,73
engominarse	verbo	120	12,33	812,86	10,95
crespo	adjetivo	113	12,33	766,08	10,63
ensortijar	verbo	138	12,28	929,61	11,74
mechón	sustantivo	404	12,04	2.649,50	20,09
atusar	verbo	68	11,97	442,20	8,24
entrecano	adjetivo	58	11,97	377,36	7,61
fasciculado	adjetivo	11	11,9	71,01	3,31
castaño	adjetivo	300	11,81	1.917,94	17,32
rizado	adjetivo	212	11,8	1.354,14	14,56
escarpia	sustantivo	11	11,79	70,07	3,31
cardado	sustantivo	14	11,68	88,15	3,74
rizado	sustantivo	11	11,61	68,69	3,31
cardar	verbo	24	11,56	149,09	4,89
cano	adjetivo	93	11,52	575,63	9,64

1 - 20 de 1521

página: 1 2 3 4 5 6 7 8 9 ... 77

Abbildung 94: CORPES – *Coapariciones* des Substantivs *pelo*.¹¹⁵⁸

Den Schülern sind nicht alle diese Lexeme bekannt, weshalb auch hier mit einem Online-Wörterbuch, beispielsweise von PONS, gearbeitet werden sollte.¹¹⁵⁹ Die Schüler betrachten die Konkordanzen der jeweiligen Kollokation, ermitteln deren Bedeutung und wählen die ihrer Meinung nach relevanten Kollokatoren aus. Ein ausführliches, wenn auch nicht vollständiges Kollokationsfeld zur Basis *pelo* könnte demnach wie folgt aussehen:¹¹⁶⁰

Kollokationsfeld: PELO**pelo** largo / corto / tieso / fino / ralo**pelo** liso / lacio / ondulado / rizado / enrulado / crespo / ensortijado / encrespado**pelo** sedoso / pajizo / erizado / hirsuto**pelo** revuelto / alborotado / despeinado / mojadouna persona de **pelo** rubio / negro / gris / castaño / entrecano / canoso / pajizo / azabache / cobrizo / rojizo**pelo** cardado / rapado / teñido [de negro/rubio]tener / llevar el **pelo** recogido / peinado en una coleta / en un moño / en trenzasllevar el **pelo** suelto¹¹⁵⁸ Real Academia Española (ed.): *CORPES. Coapariciones*. Op. cit. (31.08.2018).¹¹⁵⁹ Vgl. Pons (ed.): op. cit. (03.11.2017).¹¹⁶⁰ Das Kollokationsfeld zum Lexem *pelo* wurde in Anlehnung an Bahns Kollokationsfeld zum Lexem *car* erstellt. Vgl. Bahns, Jens (1997): *Kollokationen und Wortschatzlernen im Englischunterricht*. Tübingen: Gunter Narr, S. 21.

pelo [cortado] al rape / engominado / peinado con gomina / peinado a lo garçon
 peinar / cepillarse / alisarse / desenredar el **pelo**
 lavarse / cortarse / rapar[se] / teñir[se] / atusarse el **pelo**

Abbildung 95: Kollokationsfeld zum Substantiv ‚pelo‘.

Aufgrund der vielen bislang unbekannten Lexeme kann die deutsche Bedeutung ebenfalls in Klammern dahinter angegeben werden, um den Schülern die Nutzung des Kollokationsfelds zu erleichtern:

Kollokationsfeld: PELO

pelo largo (lang) / corto (kurz) / tieso (kräftig) / fino (fein) / ralo (licht, schütter)
pelo liso (glatt) / lacio (glatt) / ondulado (gewellt) / rizado (lockig, gelockt) / enrulado (lockig) / crespo (kraus) / ensortijado (kraus) / encrespado (kraus)
pelo sedoso (seidig) / pajizo (strohig) / erizado (struppig) / hirsuto (struppig)
pelo revuelto (durcheinander) / alborotado (verwuschelt) / despeinado (ungekämmt, zerzaust) / mojado (nass)
 una persona de **pelo** rubio (blond) / negro (schwarz) / gris (grau) / castaño (kastanienbraun) / entrecano (graumeliert) / canoso (grauhaarig) / pajizo (strohgelb) / azabache (rabenschwarz) / cobrizo (kupferfarben) / rojizo (rötlich)
pelo cardado (toupirt) / rapado (kurz geschoren) / teñido [de negro/rubio] (gefärbt)
 tener / llevar el **pelo** recogido / peinado en una coleta (Pferdeschwanz) / en un moño (Haarknoten, Dutt) / en trenzas (Zöpfe)
 llevar el **pelo** suelto (offen tragen)
pelo [cortado] al rape (kurz/kahl geschoren) / engominado (gegelt, zurückgegelt) / peinado con gomina (mit Gel frisiert) / peinado a lo garçon (Bubikopf)
 peinar (kämmen) / cepillarse (bürsten) / alisarse (auskämmen) / desenredar el **pelo** (durchkämmen)
 lavarse (waschen) / cortarse (schneiden) / rapar[se] (kahl scheren) / teñir[se] (färben) / atusarse el **pelo** (Frisur/Haare richten)

Abbildung 96: Kollokationsfeld zum Substantiv ‚pelo‘ inklusive der Angabe der deutschen Entsprechung.

Die Erstellung derartiger Kollokationsfelder ist besonders zur Vorbereitung auf die Bildbeschreibung im Hinblick auf mündliche oder auch schriftliche Schulaufgaben geeignet. Dabei können die Schüler in Paaren oder Kleingruppen zu unterschiedlichen Basen, beispielsweise zu *cara*, *ojos*, *nariz*, *piel*, *cuerpo* und *ropa*, Kollokationsfelder erstellen und diese der Klasse präsentieren. In diesem Fall ist die Anordnung der Ergebnisse nach dem Wert der MI sogar wünschenswert, da die Frequenz der Kombination hierbei weniger berücksichtigt wird. Zudem ist die Anordnung nach

Wortart nicht sinnvoll, da Ergebnisse aller Wortarten in das Kollokationsfeld eingehen sollen, denn die Beziehung zur Basis ist hier ausschlaggebend.

Allerdings muss in diesem Zusammenhang erneut darauf hingewiesen werden, dass die Ergebnisse des CORPES bei der Anordnung nach dem Wert der MI teilweise Fragen aufwerfen und die Lehrkraft die durchzuführenden Suchen daher selbst zuvor durchführen sollte. Aufgrund der Auflistung der Ergebnisse nach diesem Wert werden, wie bereits in Kapitel 4.1.1.3.1. erwähnt, Lexeme, die ausschließlich zusammen mit dem Suchbegriff auftreten, an erster Stelle genannt, auch wenn das gemeinsame Auftreten sich auf insgesamt zehn Vorkommen beschränkt. Die freie Suche der Schüler ist hier also nicht problemlos möglich. Zudem wird in der Funktion *Coapariciones* im Intervall von fünf Lexemen links und rechts des eingegebenen Schlagworts gesucht, wodurch viele Treffer angezeigt werden, unter denen sich auch andere Kombinationen und nicht nur Kollokationen befinden. Dennoch kann die Ermittlung von zusammen vorkommenden Lexemen im Abstand von fünf Einheiten für die Erstellung eines Kollokationsfelds auch von Vorteil sein, wie am Kollokationsfeld zu *pelo* gesehen werden kann, da sonst Ausdrücke wie *llevar el pelo peinado en una coleta* nicht als zusammengehörig erkannt würden (siehe Abbildungen 95 und 96).

Ein derartiges Kollokationsfeld könnte allerdings auch anhand des CdE:Nuevo (siehe Kapitel 2.6.1.) erarbeitet werden. Dazu sollen die Schüler im Reiter *Colocados* das Substantiv *pelo* als *Palabra/frase* eingeben und im Umfeld von bis zu fünf Lexemen rechts und links beispielsweise nach Adjektiven suchen (*Colocados: adj.TODOS*). Auf diese Weise kann ebenfalls nach Verben im direkten Umfeld von *pelo* gesucht werden. Ebenso ist es möglich, Lexeme unterschiedlicher Wortarten in der Umgebung der Basis auf einmal herauszuarbeiten, indem ein Asterisk im Feld *Colocados* eingetragen wird.

Auch wenn die Schüler nicht selbstständig in Korpora suchen, so kann die Lehrkraft anhand derselben ein Arbeitsblatt mit einem Kollokationsfeld erstellen, welches die Schüler zur Aneignung des Vokabulars nutzen können. Falls die Erstellung eines ganzen Kollokationsfelds zu viel Arbeit sein sollte, können auch verschiedene Gruppen an unterschiedlichen Bereichen arbeiten. Eine Gruppe könnte nach Adjektiven suchen, eine andere nach Verben. Am Ende entscheiden die Schüler, welche Ausdrücke in das Kollokationsfeld aufgenommen werden und welche nicht. Dabei können nicht nur die Treffer der Liste der *Coapariciones* ausgewählt werden, sondern auch Lexeme, die den

Schülern bei der Durchsicht der Konkordanzen ins Auge stechen. Auch der sprachliche Kontext der Lexeme spielt eine wichtige Rolle, damit die Schüler wissen, wie die Lexeme genutzt werden. Auf diese Weise finden sie heraus, dass der deutsche Ausdruck „(kurze, lange, blonde, glatte) ... Haare haben“ im Spanischen mit *tener el pelo...*, *llevar el pelo...* oder *una persona/un hombre de pelo...* versprachlicht wird. Die Lehrkraft sollte den Schülern zudem nahelegen, dass sie bei Zweifeln immer direkt in den Konkordanzen nachsehen, da in der Liste der *Coapariciones* beispielsweise das Verb *engominarse* aufgeführt wird, der Ausdruck in den Konkordanzen allerdings in der Regel *pelo engominado* lautet. Der Grund dafür ist, dass in der Liste der *Coapariciones* nur Lemmata angegeben sind und *engominado* daher als Partizip der Verbform unter der Infinitivform kategorisiert wird. Zudem können eventuelle Bedeutungsnuancen oder Kombinationsbeschränkungen, beispielsweise im Hinblick auf die in den Abbildungen 95 und 96 aufgeführten verschiedenen spanischen Begriffe für ‚glattes‘, ‚lockiges‘ oder ‚krauses‘ Haar, anhand der Konkordanzen ermittelt werden.

Das Erlernen von Kollokationen anhand dieser durch Korpora erstellter Kollokationsfelder ist vor allem deshalb sinnvoll, weil die Schüler dadurch den sprachlichen Kontext, also die typische Struktur um das betreffende Lexem herum, kennenlernen. Im Hinblick auf mündliche Prüfungen lernen Schüler für gewöhnlich viel spezielles Vokabular zur Beschreibung von Personen, wie beispielsweise das Lexem *moño*, wissen aber nicht, wie sie es in einen Satz einbetten sollen. Wenn sie sich selbstständig Kollokationen zu diesem Lexem erarbeitet haben, wissen sie, dass es normalerweise als (*llevar el pelo*) *recogido en un moño* verwendet wird.

Durch die Analyse von Kollokationen können somit sowohl die Bedeutung von Lexemen verdeutlicht und verschiedene Bedeutungsnuancen ermittelt als auch der gewöhnliche Kontext veranschaulicht werden. Einige Lexeme, die das Korpus anzeigt, werden den Schülern nicht geläufig sein, diese können sie jedoch ohne Probleme in einem Wörterbuch nachschlagen und auf diese Weise durch die intensive Beschäftigung mit dem Vokabular ihre Ausdrucksfähigkeit erweitern.

Neben dem CORPES und dem CdE:Nuevo kann auch das Korpus *Google Books: Spanish* (siehe Kapitel 2.6.2.), welches wie das CdE von Mark Davies erstellt wurde, für die Ermittlung von Kollokationen genutzt werden. Dieses Korpus hält zwar nicht alle Suchoptionen der anderen Korpora von Davies bereit, doch es ermöglicht unter anderem

die Suche nach Kollokationen. Bei der Suche nach dem Substantiv *cielo* zusammen mit einem Adjektiv ergeben sich folgende Treffer:

The screenshot shows the Google Books: Spanish search interface. On the left, the search string is 'cielo [J*]'. The POS list is set to 'adj.ALL'. The sections are set to 'SHOW'. The sorting is set to 'FREQUENCY'. The minimum frequency is set to 100. On the right, a table lists 16 search results, each with a rank, the word/phrase, the chart type (G), and the total count.

	WORD(S)	CHARTS	TOTAL
1	cielo azul	G	54379
2	cielo abierto	G	34289
3	cielo estrellado	G	16927
4	cielo gris	G	9058
5	cielo despejado	G	7720
6	cielo nocturno	G	7177
7	cielo claro	G	7123
8	cielo limpio	G	4451
9	cielo negro	G	4222
10	cielo nublado	G	3883
11	cielo oscuro	G	3614
12	cielo puro	G	3390
13	cielo límpido	G	3026
14	cielo nuevo	G	3022
15	cielo plumizo	G	2541
16	cielo luminoso	G	2231

Abbildung 97: Google Books: Spanish – Kollokationen mit dem Substantiv *cielo* und einem Adjektiv.¹¹⁶¹

Während das *Spa_newscrawl_2011*, wie in Kapitel 4.1.1.3.2. veranschaulicht, Lexeme bereitstellt, die mit dem Suchbegriff in einem mehr oder weniger weiten sprachlichen Kontext vorkommen, bietet das *Google Books: Spanish* eine Auflistung von Kollokatoren, die direkt neben dem Suchbegriff verwendet werden. Erstere können für die Erstellung einer Mindmap genutzt werden, während Kollokationen diese Mindmap erweitern oder zur Erstellung eines Kollokationsfelds herangezogen werden können, wie im vorliegenden Kapitel beschrieben. Während das CORPES jedoch Lexeme aller Wortarten abhängig von der Höhe ihres MI-Werts auflistet, wird im Korpus *Google Books: Spanish* nach einer bestimmten Wortart gesucht. Sofern im *Google Books: Spanish* mittels eines Asterisks nach allen Wortarten beziehungsweise jeglichen Zeichen gesucht wird, werden vor allem Funktionswörter und Interpunktionszeichen in der Auflistung aufgeführt. Für die Erstellung eines Kollokationsfelds müssten also Suchen nach unterschiedlichen Wortarten durchgeführt werden, dafür können die Treffer in der jeweiligen Wortart allerdings umgehend verglichen werden.

¹¹⁶¹ Davies: *Google Books: Spanish*. op cit. (20.01.2018).

4.1.1.3.4. Erarbeitung von Nomen-Adjektiv-Kollokationen

Die Suche nach Kollokationen ist besonders interessant im Bereich der Kombination von Substantiven und Adjektiven. Um ihre Ausdrucksfähigkeit zu verbessern, können die Schüler in einem Korpus Adjektive zu vorgegebenen Substantiven suchen, beispielsweise für das häufig gebrauchte Substantiv *problema*. Sofern man im CdE:Orig (siehe Kapitel 2.1.1.3.) nach [problema] mit einem beliebigen Adjektiv sucht, ergeben sich neben Adjektiven zur genaueren Bezeichnung der Probleme (*problemas sociales/ económicos/ políticos/ internos*) auch solche zur bloßen Verstärkung, welche von Schülern zur Verbesserung ihrer Textproduktionen verwendet werden können, beispielsweise *problema grave/ principal/ fundamental/ importante/ serio/ básico*.

Ein Arbeitsblatt, mithilfe dessen die Schüler selbstständig in einem Korpus nach passenden Adjektiven für besonders frequente Substantive suchen können, scheint in diesem Zusammenhang sinnvoll. Einen Vorschlag hierfür bietet das Arbeitsblatt 4 – *Adjetivos adecuados* (Anhang 13). Das CdE ist für diese Übung am besten geeignet, da hier speziell nach Substantiv-Adjektiv-Verbindungen gesucht werden kann. Um passende Adjektive zu ermitteln, soll der Schüler im CdE:Orig (Version 2008) das jeweils vorgegebene Substantiv in eckigen Klammern eingeben, beispielsweise [ventaja], um nach Treffern im Singular und Plural zu suchen. Außerdem ist die Sucheinstellung so anzupassen, dass alle Adjektive im Umkreis von einem Lexem links und rechts des Suchbegriffs ermittelt werden. Daraufhin sollen einige Adjektive, die für den Schüler passend scheinen, in der zweiten Spalte vermerkt und ein Beispielsatz in der dritten Spalte angegeben werden, um sicherzustellen, dass die Schüler in der Lage sind, das betreffende Adjektiv korrekt zu verwenden. Zum Verfassen des Beispielsatzes können die Schüler sich an die Konkordanzen des Korpus anlehnen sowie ein Online-Wörterbuch zurate ziehen.

Ziel dieser Übung ist es, verschiedene bekannte und neue Adjektive kennenzulernen, welche von Muttersprachlern häufig in Kombination mit dem vorgegebenen Substantiv verwendet werden, um die Ausdrucksfähigkeit zu erweitern. Jedes Adjektiv bringt dabei eine unterschiedliche Bedeutungskomponente mit sich, die es zu berücksichtigen gilt, doch die Konkordanzen helfen den Schülern, die Kombination im jeweiligen sprachlichen Kontext zu sehen, zu verstehen und in einem Beispielsatz

anzuwenden. Vor Beginn der Übung sollte mit der Klasse vereinbart werden, wie viele Adjektive aufzulisten sind.

Nach der Erstellung einer derartigen Liste wie in Arbeitsblatt 4 ist es ratsam, die Schüler einen Text schreiben zu lassen, in dem sie eine Mindestanzahl an zuvor gesuchten Ausdrücken verwenden (beispielsweise mindestens zehn Kombinationen). Sofern die Schüler im Umgang mit Computern beziehungsweise Korpora zu unerfahren sind oder kein Computerraum zur Verfügung steht, ist es auch denkbar, dass der Lehrer die Liste passender Adjektive selbst erstellt und sie den Schülern, eventuell sogar mit ein paar Konkordanzen, zur Verfügung stellt.

Diese Übung ist ab Anfang des dritten Lernjahrs geeignet und dient vor allem der Verbesserung der Ausdrucksfähigkeit, um den Schülern zu mehr Abwechslung in ihren Textproduktionen zu verhelfen. Eine ähnliche Aufgabe zur Erweiterung der Wortschatzkenntnisse, allerdings im Hinblick auf die Ermittlung von Synonymen für häufig verwendete Adjektive, findet sich in Arbeitsblatt 6 (siehe Kapitel 4.1.3.4., Anhang 15), welches ebenso ab Anfang des dritten Lernjahrs angewandt werden kann. Arbeitsblatt 6 bietet eine Art Kombination zwischen der Suche nach Synonymen und derjenigen nach Kollokationen, da nach passenden Adjektiven in einer speziellen Bedeutung für bestimmte Substantive gesucht wird, beispielsweise für Synonyme von *grande* zum Substantiv *problema*.

In anderen Korpora, beispielsweise dem CORPES der *Real Academia Española* (siehe Kapitel 2.1.1.2.) kann ebenso nach Adjektiven gesucht werden, die zusammen mit einem bestimmten Substantiv verwendet werden, allerdings ist das CdE hier übersichtlicher. Obwohl das CORPES prinzipiell für Schulklassen gut geeignet ist, aufgrund der Tatsache, dass man es ohne Registrierung nutzen kann, sollte bei diesem Arbeitsblatt dennoch auf das CdE zurückgegriffen werden, da hier auf einen Blick die häufigsten Adjektive aufgelistet werden. Im CORPES können lediglich passende Beispiele als Konkordanzen angezeigt werden, ohne dass jedoch Informationen über deren Häufigkeit angegeben werden. Während der Nutzer im CORPES alle gefundenen Konkordanzen durchsehen und schließlich eine eigene Suchanfrage für jede interessante Kombination starten muss, bietet das CdE bei jeder Suche zunächst eine Übersicht über die gefundenen Lemmata inklusive deren Frequenz, aus denen im Folgenden diejenigen

4. Nutzung von Korpora in der fremdsprachlichen Unterrichtspraxis

von Interesse ausgewählt werden können, um sich die betreffenden Konkordanzen anzeigen zu lassen.

Soll dennoch mit dem CORPES gearbeitet werden, so wird mittels der Suchoption *Proximidad* speziell nach bestimmten Wortarten in der Umgebung des Suchbegriffs gesucht, beispielsweise nach Adjektiven im Umfeld von einem Lexem links und rechts des Substantivs *problema*.

The screenshot shows the CORPES search interface. At the top, there are tabs: 'Concordancias', 'Coapariciones', 'Configuración', 'Ayuda', 'Modo de cita', and 'Sugerencias'. The 'Concordancias' tab is active. Below the tabs, there are search filters: 'Lema' (problema), 'Forma' (empty), 'Clase de palabra' (sustantivo), 'Grafía original' (checked), and '+ Subcorpus'. Below these, there are more filters: 'Número' (empty), 'Género' (empty), 'Tipo' (empty), and 'Proximidad' (Limpiar). The 'Proximidad' section has a sub-section with 'Lema' (empty), 'Forma' (empty), 'Clase de palabra' (adjetivo), 'Distancia' (radio button), 'Intervalo' (radio button), and 'Grado' (empty). Below this, there are buttons: '+', '-', 'Concordancia', 'Estadística', and 'Nueva consulta'. The results section shows '31.764 casos en 18.484 documentos.' and a table with columns: 'REF. (Clasificación, país)', 'CONCORDANCIA', and 'Ordenar'. The table lists 20 results, each with a reference number, a country, a snippet of text, and the word 'problema' or 'problemas' highlighted in blue. The results are sorted by 'Ordenar'.

REF. (Clasificación, país)	CONCORDANCIA	Ordenar
1 2001 Arg.	No hay ningún problema	personal. Pero ya
2 2001 Cuba	Yo creo que los principales problemas	de Cuba se deber
3 2001 Arg.	36. ¿Qué opina del problema	vasco? ¿Votaría al
4 2001 Esp.	63. Pedro te noto mala cara por la videocámara, has estado enfermo o tienes algún problema	gordo. Contamina
5 2001 Esp.	30. es un disco sobre los problemas	sociales y actuales
6 2001 Esp.	una exploración amena y documentada, que toca todas las cuerdas y no oculta ningún problema	filosófico, moral o
7 2001 Méx.	tener que soportar su presencia, su conversación, sus ambiciones mezquinas, sus problemas	cotidianos, su car
8 2001 Méx.	-El problema	mayor de esta reg
9 2001 Méx.	-Tengo un problema	táctico, rigidez inv.
10 2001 Chile	el marco vacío donde debiera refulgir la sonrisa del príncipe. Tus amigos son un problema	elevado al cubo: s
11 2001 Chile	me parecieron entonces, y por primera vez desde que había iniciado el viaje, mis problemas	sentimentales del
12 2001 Chile	podís mandar cambiar. No estoy para pasar malos ratos contigo, tengo suficientes problemas	...
13 2001 Chile	iconos massmediáticos, no mediante intercambios de información o el análisis de los problemas	populares".
14 2001 Col.	PROBLEMAS	TÉCNICOS
15 2001 Ven.	Estados Unidos. Quedaban en pie los hombres de Matos que, a estas alturas, eran ya un problema	menor. En julio, el
16 2001 Ven.	cuarenta y dos años de ausencia volvió a Venezuela con el propósito de resolver algunos problemas	personales. Venía
17 2001 Ven.	político hizo todo lo posible por dar respuesta, de la manera eficaz, a los grandes problemas	nacionales. Se ati
18 2001 Ven.	Cabré. Si en aquellos lejanos días -dijo una vez Mariano Picón Salas- los grandes problemas	nacionales podían
19 2001 Ven.	presenta a esta región como de fácil acceso, cercano a los centros urbanos, sin graves problemas	de comunicación,
20 2001 Ven.	consecuencias inesperadas que colocaría a las autoridades españolas frente a un serio problema	de lealtad de los h

Abbildung 98: CORPES – Suche nach Adjektiven im Umfeld von *problema*.¹¹⁶²

Des Weiteren ist es im CORPES möglich, über die verschiedenen Eingabezeilen nach miteinander kombinierten Wortarten zu suchen, beispielsweise nach Substantiven mit zwei Adjektiven. Diese Einstellung ist über die Angabe der *Distancia* auszuwählen.

4.1.1.3.5. Integration von neuem Vokabular in einen größeren sprachlichen Kontext

Im Hinblick auf die verschiedenen Verfahren zur Ermittlung von Kollokationen ist es auch denkbar, die traditionelle Führung des Vokabelhefts abzuändern und um Kollokationen sowie reale Beispielsätze zu ergänzen. Bislang führen Fremdsprachenlerner ein zwei- oder dreispaltiges Vokabelheft, wobei in der linken Spalte das fremdsprachliche Lexem, in der rechten Spalte die deutsche Entsprechung und in der

¹¹⁶² Real Academia Española (ed.): *CORPES. Entrada Compleja*. Op. cit. (31.08.2018).

mittleren Spalte ein Beispielsatz, eine Erklärung, ein Synonym, ein Antonym oder eine Zeichnung steht, je nachdem, was in der Mittelspalte des Lehrbuchs aufgeführt wird. In manchen Lehrwerken gibt es allerdings keine derartige Mittelspalte beziehungsweise bei Lehrwerken in höheren Jahrgangsstufen ist teilweise wieder eine zweiseitige Auflistung des Vokabulars zu finden. Bei fortgeschrittenen Schülern könnte daher die Mittelspalte im Vokabelheft von den Schülern selbst mittels eines Korpus ausgefüllt werden. So könnten die Lerner als Hausaufgabe eine passende Kollokation oder einen Beispielsatz mit dem gewöhnlichen Kontext des jeweiligen Lexems in die Mittelspalte eintragen. Durch die intensive Beschäftigung mit dem Vokabular und dessen sprachlicher Umgebung werden die Lexeme besser in das bereits vorhandene Wissen integriert und somit wird ein höherer Prozentsatz im Gedächtnis behalten. Ist beispielsweise das Substantiv *crimen* Teil des neu zu erlernenden Wortschatzes, könnten die Schüler ein passendes Verb dazu suchen (zum Beispiel *cometer un crimen*) oder einen Beispielsatz dazu schreiben.

Eine andere Option wäre, dass die Schüler nach der Einführung neuer Vokabeln oder auch sofern sie bestimmte Lexeme für die Textproduktion benötigen, zu Hause einige Verwendungsweisen, also den gewöhnlichen sprachlichen Kontext und somit auch Kollokationen, mithilfe eines Korpus herausfinden und diese in der nächsten Unterrichtsstunde vorstellen. Beispiele für Kollokationen, deren sprachlichen Kontext die Schüler zu Hause ermitteln könnten, sind Ausdrücke wie *hacer una encuesta*. Sofern Schüler in ihren eigenen Sprachproduktionen das Lexem *encuesta* verwenden möchten, aber nicht wissen, wie der Ausdruck „eine Umfrage durchführen“ ins Spanische zu übertragen ist, können sie das Substantiv *encuesta* als Suchausdruck im CdE:Orig (siehe Kapitel 2.1.1.3.) eingeben und mittels *verb.TODOS* im Abstand von beispielsweise vier Lexemen vor dem Suchausdruck nach dem Kollokator suchen. Wenn sie die Verben durchsehen, finden sie eine hohe Trefferzahl für sämtliche Formen des Verbs *hacer*, aber auch das Verb *realizar* in der gleichen Umgebung. Des Weiteren werden ihnen die Treffer *reveló* und *muestra* angezeigt. Diese letzten beiden Verben können zwar nicht an die Stelle von *hacer* treten, aber es ist möglich, sie bei einer Änderung des Satzes von Aktiv auf Passiv in ähnlichen Zusammenhängen zu verwenden. Mithilfe der Angabe dieser Verben können die Schüler selbstständig verschiedene Sätze zum selben Thema formulieren und bei Unklarheiten zur Integration der Verben in den Satz die jeweiligen

Beispiele im Korpus zurate ziehen. Die Schüler müssen bei der Analyse der Suchergebnisse immer den Gebrauch der Lexeme im Kontext der Beispiele betrachten, um entscheiden zu können, ob sie diese wie gewollt in den Text einbringen können, dafür bieten sich ihnen allerdings auch mögliche Strukturen an, welche sie in ihren eigenen Produktionen gebrauchen können. Einige dieser Kollokationen, so auch *hacer una encuesta*, werden bei Online-Wörterbüchern durchaus im jeweiligen Eintrag (*encuesta*) angezeigt, jedoch finden die Schüler hier nur einen Ausdruck. Im Online-Wörterbuch Deutsch-Spanisch von PONS wird im Eintrag von *encuesta* zum Beispiel nur *hacer una encuesta* angezeigt, mehr allerdings nicht.¹¹⁶³ Das Korpus bietet ihnen verschiedene Ausdrucksweisen sowie Formulierungshilfen.

Bei der Suche nach Kollokationen in Korpora ist allerdings auch die An- oder Abwesenheit des bestimmten oder unbestimmten Artikels für den korrekten Sprachgebrauch von Bedeutung. Natürlich wollen Lerner einer Fremdsprache zunächst wissen, welches Verb mit einem bestimmten Substantiv verwendet wird. Bei der Analyse der Konkordanzen wird jedoch auch klar, ob zwischen den beiden gesuchten Ausdrücken gewöhnlich ein Artikel steht und ob dieser häufiger definit oder indefinit ist. So wird beispielsweise kein Artikel gesetzt bei *darse cuenta de*, bei *darse el gusto de* benötigt man allerdings den bestimmten Artikel. Auch im Fall des teilweise für deutsche Schüler verwirrenden Artikelgebrauchs im Hinblick auf Ländernamen wie *la Argentina* oder *el Perú* kann ein Korpus von Nutzen sein, denn die Schüler können anhand der Konkordanzen zum jeweiligen Substantiv sofort erkennen, ob und wann ein bestimmter Artikel gebraucht wird.

Ein weiteres Beispiel ist das Lexem *investigación*. Hier finden die Schüler die Verben *hacer*, *realizar*, *iniciar* und *abrir* im CdE:Orig. Sie können je nach Wahl des Verbs in den Konkordanzen nachsehen, wie das Substantiv gewöhnlich mit diesen Verben verbunden wird, und finden so heraus, dass zwischen *hacer / realizar / abrir* und *investigación* sowohl der unbestimmte als auch der bestimmte Artikel stehen kann und dass auch *hacer / realizar* und *abrir investigación* ohne Artikel verwendet werden können, wobei dies selten der Fall ist. Zwischen *iniciar* und *investigación* steht bei den Beispielen im Korpus allerdings immer ein Artikel, meist der unbestimmte, aber es finden sich auch Beispiele mit dem bestimmten Artikel oder dem Possessivbegleiter. Es muss

¹¹⁶³ Vgl. Pons (ed.): op. cit. (10.08.2015).

jedoch hinzugefügt werden, dass für eine Analyse der Konkordanzen eines Korpus ein gewisses Maß an Sprachkenntnis erforderlich ist. Schüler sollten also frühestens am Ende des dritten Lernjahrs selbst Konkordanzen analysieren und nach bestimmten Strukturen suchen.

Häufig kann beobachtet werden, dass ein Lexem im Großteil der Fälle mit einem bestimmten anderen Lexem gebraucht wird und man es daher selten außerhalb dieser Kombination antrifft. Dieser restriktive Gebrauch wird in Wörterbüchern nicht so ersichtlich wie in den Konkordanzen eines Korpus. Durch das Erlernen dieser Zusammenhänge werden die Schüler sprachlich versierter und verbessern ihre kommunikative Kompetenz, da sie zunehmend größere Einheiten in immer kürzerer Zeit produzieren können. Mittels der Möglichkeit, nach Kollokationen zu suchen, erleichtert das Korpus die Suche nach relevanten Wortverbindungen, da die frequentesten Verbindungen in abfallender Reihenfolge aufgelistet werden. Nun müssen die einzelnen Konkordanzen nur noch aufgerufen und analysiert werden. Zudem ist es möglich, anhand der Konkordanzen positive oder negative Konnotationen zu ermitteln.

Des Weiteren können die Schüler im Falle von Kollokationen mit ähnlicher Bedeutung, beispielsweise bei *error garrafal* / *craso* / *grave* mithilfe des Korpus herausfinden, welcher der Ausdrücke häufiger vorkommt beziehungsweise signifikanter ist, und auch, in welchen Bereichen die Lexeme verwendet werden, also ob sie zu einem bestimmten Register gehören, ob sie eher mündlich oder schriftlich, in akademischen oder literarischen Texten gebraucht werden. Weitere Informationen zur Suche nach Synonymen sind in Kapitel 4.1.3. zu finden.

4.1.2. Phraseologismen

Im *Lexikon der Sprachwissenschaft* werden Phraseologismen als Phraseme bezeichnet, es wird jedoch zudem angegeben, dass diese auch unter den Begriffen Syntagma, Idiom, idiomatische Wendung, Makrosemem, Redewendung oder eben Phraseologismus zu finden sind. Ein Phrasem wird dort wie folgt definiert:

„Feste, mehrgliedrige Wortgruppe bzw. Lexikoneinheit mit folgenden Eigenschaften: a) die Gesamtbedeutung kann nicht aus der Bedeutung der Einzelelemente abgeleitet werden, vgl. *jemanden auf die Palme bringen* >jemanden wütend machen<; b) der Austausch von Einzelelementen ergibt (anders als bei nichtphraseologischen Syntagmen) keine systematische Bedeutungsveränderung, vgl. *jemanden auf die Birke bringen*, c) in wortwörtlicher Lesart ergibt sich eine homophone (= gleich lautende), nicht phraseologische Variante, für die die Bedingungen

4. Nutzung von Korpora in der fremdsprachlichen Unterrichtspraxis

a) und b) nicht gelten; [...] Zwischen wortwörtlicher Lesart und phraseologischer Lesart besteht diachronisch gesehen häufig ein Zusammenhang [...]. In diesen Fällen ist die Behandlung des P. als unanalysierbare Lexikoneinheit unzureichend.“¹¹⁶⁴

Lemnitzer/Zinsmeister bezeichnen Phraseologismen ebenfalls als Phraseme und schreiben ihnen drei Merkmale zu: Erstens sind sie nicht immer wohl geformt in Bezug auf die Grammatik der entsprechenden Sprache. Zweitens sind sie semantisch intransparent, die einzelnen Bestandteile haben demnach nicht die gleiche Bedeutung wie in freier Verwendung. Schließlich sind sie nur begrenzt modifizierbar, beispielsweise durch die Ergänzung von Adjektiven.¹¹⁶⁵

In dieser Arbeit soll in Anlehnung an obige Definitionen für feste, mehrgliedrige Gruppen von Lexemen, deren Gesamtbedeutung nicht aus den Einzelementen abgeleitet werden kann, bei denen sich durch den Austausch von Einzelementen keine systematische Bedeutungsveränderung ergibt und die nur begrenzt modifiziert werden können, der Begriff ‚Phraseologismus‘ verwendet werden.

Auf die Abgrenzung des Begriffs ‚Phraseologismus‘ von demjenigen der Kollokation sowie die Bedeutung der Beherrschung von Phraseologismen für den Sprecher einer Sprache ist bereits in Kapitel 4.1.1. eingegangen worden. Im Folgenden soll daher lediglich die Erweiterung der Ausdrucksfähigkeit anhand der Erarbeitung von Phraseologismen veranschaulicht werden.

Aufgrund der semantischen Intransparenz, die zumindest ein Element des Phraseologismus betrifft, sind Phraseologismen für Fremdsprachenlerner schwer zu verstehen und noch komplexer in der eigenen Anwendung. Die begrenzte Modifizierbarkeit stellt den Nutzer zusätzlich vor Herausforderungen. Aus diesem Grund sollte im Fremdsprachenunterricht nicht nur auf Kollokationen, sondern auch auf Phraseologismen eingegangen werden.

Wie bei den Kollokationen gibt es auch bei den Phraseologismen Schwierigkeiten, diese in Wörterbüchern zu finden. So werden Ausdrücke wie *estar como un tren* meist unter dem Lemma *tren* geführt, wo sie allerdings nur bei der Textrezeption, nicht aber bei der Textproduktion gefunden werden können. Laut Pöll werden Phraseologismen

¹¹⁶⁴ Bußmann (2008): op. cit., S. 530. In Anlehnung an: Kispál, Tamás (2007): „Sprichwörtersammlungen“. In: Burger, H./ Dobrovolskij, D./ Kühn, P./ Norrick, N. R. (2007): *Phraseologie. Ein internationales Handbuch der zeitgenössischen Forschung*. HSK Band 28.1. Berlin [u.a.]: De Gruyter, S. 414-423.

¹¹⁶⁵ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 145-146.

häufig unter dem ersten Autosemantikon geführt, was bei *hablar por los codos* das Verb *hablar* wäre. Teilweise werden sie allerdings auch unter dem ersten Substantiv aufgelistet und daraufhin wird in einer festgelegten Reihenfolge auf die anderen Wortarten zurückgegriffen. Zudem gibt er an, dass Phraseologismen teilweise falsch zitiert werden, indem die phraseologische Wortart verändert wird. So wird der Ausdruck *de bote en bote* in Wörterbüchern bisweilen unkorrekterweise als *estar de bote en bote* ‚brechend voll sein‘ geführt, da er häufig mit *estar* vorkommt. Ebenso problematisch sieht er die Doppelnennung von Elementen (*irsele a alg. la cabeza* ‚schwindelig werden‘) sowie die verschiedenen Varianten von Phraseologismen (*anda y vete a pasear/ al carajo/ al diablo* ‚scher dich zum Teufel‘). Weiter vertritt Pöll die Meinung, dass pragmatisch-kommunikative Gebrauchsrestriktionen in Wörterbucheinträgen nicht immer klar ersichtlich sind. So werden *echar chispas* (‚sich erhitzen, ärgern‘) oder *brillar por su ausencia* (‚durch Abwesenheit glänzen‘) meist in der dritten Person gebraucht, wobei Letzteres im Spanischen ebenso gut eine Sache zum Subjekt haben kann. Diese Restriktionen müssten idealerweise mithilfe eines Beispiels illustriert werden.¹¹⁶⁶ Die erste Schwierigkeit besteht also darin, die Phraseologismen im Wörterbuch zu finden, und die zweite darin, sie korrekt in den sprachlichen Zusammenhang einzubauen. Aus diesem Grund ist auch hier, wie bei den Kollokationen, die Nutzung von Korpora empfehlenswert, damit die Schüler nicht nur den jeweiligen Phraseologismus und dessen Bedeutung, sondern auch dessen Einordnung in den sprachlichen Kontext kennenlernen, um diesen korrekt verwenden zu können.

Ein Beispiel für die Erarbeitung von Phraseologismen anhand von Korpora bietet Arbeitsblatt 5 *Dichos españoles* (Anhang 14). Dieses Arbeitsblatt soll mit der alten Version des CREA bearbeitet werden (siehe Kapitel 2.1.1.1.). Das Korpus CREA (Version 2015) ist nicht annotiert und bietet daher die Möglichkeit, ganze Wortsequenzen in die Suchmaske einzugeben, ohne den Suchausdruck in einzelne Lexeme unterteilen und in unterschiedliche Suchfelder eintragen zu müssen, wie dies bei der annotierten Version des CREA und dem CORPES (siehe Kapitel 2.1.1.2.) der Fall ist. Zudem können im CREA (Version 2015) nicht nur Konkordanzen, sondern auch ganze Textausschnitte angezeigt werden. Durch einen Klick auf den jeweiligen Treffer wird die betreffende Textstelle mit einigen Sätzen vor und nach dem Suchbegriff angezeigt. Diese Möglichkeit

¹¹⁶⁶ Vgl. Pöll (2002): op. cit., S. 138-140.

4. Nutzung von Korpora in der fremdsprachlichen Unterrichtspraxis

erlaubt es dem Nutzer, den weiteren Kontext des Suchausdrucks zu analysieren, um dessen Verwendung genauer bestimmen zu können. Eben diese Funktion ist bei der Arbeit mit Phraseologismen von großem Nutzen.

Die Mehrheit der Phraseologismen auf Arbeitsblatt 5 könnte auch mit einem Wörterbuch ermittelt werden, dort wird allerdings nur die jeweilige deutsche Entsprechung angegeben. Durch die aktive Suche im Korpus sowie die Analyse der Konkordanzen und Textausschnitte erfahren die Schüler jedoch, in welchem sprachlichen Kontext die Phraseologismen verwendet werden. Ein Ausdruck auf dem Arbeitsblatt ist *hablar por los codos*, welcher beispielsweise auch im Online-Wörterbuch von PONS zu finden ist. Bei der Durchsicht der Konkordanzen zum Suchausdruck *por los codos* sticht dem Nutzer jedoch sofort ins Auge, dass in der Mehrzahl der Fälle *hablar hasta por los codos* verwendet wird, was im Wörterbucheintrag nicht ersichtlich ist. Auf diesem Wissen um den sprachlichen Kontext baut auch die zweite Teilaufgabe auf, bei der die Schüler Anwendungsbeispiele für die Phraseologismen angeben sollen, welche sie anhand der Konkordanzen erstellen können. Es werden zwar lediglich fünf Sätze mit korrekt verwendeten Phraseologismen von den Schülern gefordert, doch auf dem Lösungsblatt ist je ein Beispielsatz pro Ausdruck angegeben, um den typischen Kontext für jeden Phraseologismus zu veranschaulichen. Alle Beispielsätze wurden aus dem CREA (Version 2015) übernommen und lediglich gekürzt oder leicht abgeändert.

Bei den Aufgaben zu Kollokationen wurde bereits erwähnt, dass derartige Übungen für fortgeschrittene Lerner gedacht sind, da die Beherrschung derselben und vor allem ihre Anwendung in passenden Kontexten eines gewissen Maßes an Wissen über die Fremdsprache bedürfen. Für den Umgang mit Phraseologismen sollten die Schüler sogar noch weiter fortgeschritten sein. Zum einen, da die Bedeutung derselben nicht aus den Einzelkomponenten zu erschließen ist, und zum anderen, da sie in den meisten Fällen aus der Umgangssprache stammen und bei ihrem Gebrauch daher nicht nur der sprachliche, sondern auch der außersprachliche Kontext zu bedenken ist. Für fortgeschrittene Lerner gilt das Erlernen und der Gebrauch von Phraseologismen nicht nur als Maßstab für ihre Sprachkenntnisse, sondern ist zugleich von besonderem Interesse, da sie damit außersprachliche Sachverhalte auf eine typische Weise versprachlichen. Die Arbeit mit Phraseologismen motiviert die Lerner meist besonders, da sie Ausdrücke erlernen, die sie im Deutschen auch im Umgang mit ihren Freunden gebrauchen und in der Fremdsprache

bis dahin noch nicht kannten. Diesem Umstand ist es jedoch geschuldet, dass jüngere Lerner die Phraseologismen nicht nur grammatikalisch, sondern auch kontextuell falsch sowie zu häufig gebrauchen könnten und durch das Erlernen von Ausdrücken wie *¡Vete a tomar por culo!* mehr Schaden angerichtet als sprachliches Wissen vermittelt würde. Aus diesen Gründen sollte Arbeitsblatt 5, abhängig von der jeweiligen Klasse, erst ab Ende des dritten Lernjahrs, besser jedoch ab dem vierten Lernjahr, eingesetzt werden. Zudem gilt es zu beachten, dass Kollokationen generell viel häufiger gebraucht werden und sehr viel leichter für Schüler in Korpora zu ermitteln sind. Aus diesem Grund sollte den Schülern zunächst die Erarbeitung und Verwendung von Kollokationen nähergebracht werden, bevor zur Arbeit mit Phraseologismen übergegangen wird. Sofern die Schüler jedoch ein gewisses Alter beziehungsweise ein gewisses sprachliches Niveau erreicht haben, spricht nichts gegen den Einsatz dieses Arbeitsblatts, da aus Erfahrung gesagt werden kann, dass es den Schülern leicht fällt, die Bestandteile von Phraseologismen anhand eines Korpus zu erarbeiten. Nach wenigen gemeinsamen Suchen ist ihnen die Vorgehensweise bekannt und sie können problemlos weitere Phraseologismen zusammensetzen.

4.1.3. Synonyme

Bei Synonymen ist die Verwendung eines Korpus zur Erarbeitung und Festigung des Vokabulars ebenfalls zu empfehlen. In Wörterbüchern werden viele Lexeme durch die Angabe eines oder mehrerer bedeutungsähnlicher Lexeme definiert. Beim Lerner erweckt eine solche Aufzählung lexikalischer Einheiten mit ähnlicher Bedeutung oft den Eindruck, sie könnten in jeglichem Kontext gegeneinander ausgetauscht werden. Tatsächlich gibt es jedoch aufgrund der Sprachökonomie kaum Paare von Lexemen, die vollständig synonym sind und in der absolut gleichen sprachlichen Umgebung verwendet werden können.¹¹⁶⁷

Laut Pöll beschreibt der Terminus „Synonymie“ die Tatsache, dass „in einer Sprache zwei verschiedene *signifiants* mit identischem *signifié* existieren“, ob es Synonymie tatsächlich gibt, zählt seinen Angaben zufolge allerdings „zu den alten Polemiken der Sprachwissenschaft“.¹¹⁶⁸ Er gibt an, dass zwar die Effizienz der Sprache als System durch eine eindeutige Beziehung zwischen *signifiant* und *signifié* garantiert

¹¹⁶⁷ Vgl. Pöll (2002): op. cit., S. 72.

¹¹⁶⁸ Pöll (2002): op. cit., S. 72.

wird, die reale Kommunikation aber von bedeutungsgleichen oder zumindest bedeutungsähnlichen Lexemen profitiert. Für ihn sind die Verben *alquilar* und *arrendar* ein Musterbeispiel für absolute Synonymie, das heißt, dass sie in einer bestimmten Umgebung austauschbar sind und weder denotative noch konnotative Unterschiede bestehen. Außerdem sieht er *puerco* und *cerdo*, *comenzar* und *empezar* sowie *alhucema* und *lavanda* als Annäherungen an absolute Synonyme.¹¹⁶⁹

Neben der Tatsache, dass aufgrund des Ökonomieprinzips eigentlich kaum absolute Synonyme existieren, gibt Pöll allerdings an, man könne durchaus auf der Ebene der *parole* von Synonymie sprechen, da die Referenz hier der Ausgangspunkt der Analyse sein könne und daher im Sprachsystem gegebene semantische Merkmale nicht zum Tragen kommen müssten – in der *langue* hingegen gebe es nahezu keine vollkommenen Synonyme. Die Adverbien *posiblemente*, *acaso*, *probablemente*, *tal vez* und *quizá* beispielsweise stellen keine Synonyme in der *langue* dar, da sie unterschiedliche Wahrscheinlichkeitsgrade ausdrücken. In der Alltagskommunikation werden sie dennoch als Synonyme verwendet, um Wiederholungen zu vermeiden.¹¹⁷⁰

Das Phänomen der Synonymie an sich, die Frage nach der Existenz absoluter Synonymie und die Unterscheidung in unterschiedliche Arten von Synonymie wurden in der Literatur bereits hinlänglich diskutiert, beispielsweise bei Gauger (1972), Schirn (1975) und Bickmann (1978), oder in neuerer Zeit bei Kloudová (2015), Hümmer (2009) und Marková (2012), wobei die letzten beiden zur Untermauerung ihrer Theorien sogar korpuslinguistische Studien durchführten, unter anderem anhand des Deutschen Referenzkorpus.¹¹⁷¹ Obwohl die Frage nach der Existenz absoluter Synonymie unter anderem aufgrund des Widerspruchs zum Ökonomieprinzip der Sprache offen bleibt, scheint die Behandlung von Synonymen im Rahmen der Nutzung linguistischer Korpora doch vielversprechend – so kann im CdE (siehe Kapitel 2.1.1.3., 2.2.1.3. und 2.6.1.) sogar

¹¹⁶⁹ Vgl. Pöll (2002): op. cit., S. 73-74.

¹¹⁷⁰ Vgl. Pöll (2002): op. cit., S. 73.

¹¹⁷¹ Vgl. Gauger, Hans-Martin (1972): Zum Problem der Synonyme. Tübingen: Präzis. / Schirn, Matthias (1975): *Identität und Synonymie. Logisch-semantische Untersuchungen unter Berücksichtigung der sprachlichen Verständigungspraxis*. Stuttgart/ Bad Cannstadt: Frommann/ Holzboog. / Bickmann, Hans-Jürgen (1978): *Synonymie und Sprachverwendung: Verfahren zur Ermittlung von Synonymenklassen als kontextbeschränkten Äquivalenzklassen*. Tübingen: Niemeyer. / Kloudová, Věra (2015): *Synonymie und Antonymie*. Heidelberg: Universitätsverlag Winter. / Hümmer, Christiane (2009): *Synonymie bei phraseologischen Einheiten: eine korpusbasierte Untersuchung*. Frankfurt am Main [u.a.]: Lang. / Marková, Věra (2012): *Synonyme unter dem Mikroskop: eine korpuslinguistische Studie*. Tübingen: Narr.

direkt nach Synonymen gesucht werden, wobei es die dargebotenen Ergebnisse genau zu untersuchen gilt.

Im Hinblick auf die Nutzung des CdE und anderer Korpora soll in der vorliegenden Arbeit der Begriff Synonymie allerdings nicht auf bedeutungsgleiche, sondern lediglich auf bedeutungsähnliche und infolgedessen in bestimmten sprachlichen Kontexten austauschbare Lexeme angewandt werden, also im Sinne eines Synonymwörterbuchs. Hümmer (2009) bezeichnet diesen Ansatz als ‚Synonymie als Ähnlichkeitsrelation‘ im Gegensatz zur ‚Synonymie als Gleichheitsrelation‘. Nach dieser Ansicht werden Synonyme als bedeutungsähnliche lexikalische Einheiten definiert, „weil die Forderung nach Gleichheit in der Bedeutung als zu restriktiv empfunden wird“. Nach dieser Definition können Synonyme auch kleine denotative Unterschiede aufweisen, wie deutsch *schreien* – *rufen* – *brüllen*.¹¹⁷² Auch im Duden Synonymwörterbuch heißt es, vollständige Synonymie liege kaum vor, sie sei vielmehr gradierbar in Abhängigkeit von Situation und Kontext.¹¹⁷³ Bei Paaren wie *Obst* und *Früchte*, *eng* und *schmal* oder *bekommen* und *erhalten*, welche sich nicht in allen sprachlichen Kontexten austauschen ließen, sei vielmehr von Teilsynonymie oder eben partieller Synonymie zu sprechen, doch lasse sich „der Terminus Synonymie durchaus auch ohne einschränkenden Zusatz auf die zitierten Beispiele beziehen, solange deutlich bleibt, dass Synonymie keine Erscheinung des Entweder-oder, sondern eine des Mehr oder Weniger ist“.¹¹⁷⁴

Im Hinblick darauf sind Synonyme vielmehr sinnverwandte Einheiten als Lexeme, die sich durch absolute Bedeutungsgleichheit auszeichnen. Es geht daher um die Verwendung von bedeutungsähnlichen Einheiten, wie sie bisweilen in Wortfeldern dargestellt werden. Laut Duden handelt es sich bei einem Wortfeld um eine „Gruppe von Wörtern, die inhaltlich eng benachbart bzw. sinnverwandt sind“.¹¹⁷⁵ So umfasst das Wortfeld zum Thema ‚Fortbewegung auf der Erde mittels der Beine‘ unter anderem die Teilsynonyme *gehen*, *laufen*, *rennen*, *wetzen*, *schreiten*, *stolzieren*, *marschieren*, *tippeln*, *tappen* und *schlurfen*, welche in unterschiedlichem Maße austauschbar sind¹¹⁷⁶, und es

¹¹⁷² Hümmer (2009): op. cit., S. 34-35.

¹¹⁷³ Vgl. Dudenredaktion (ed.2014): *Duden – Das Synonymwörterbuch. Ein Wörterbuch sinnverwandter Wörter*. Berlin: Dudenverlag, S. 21.

¹¹⁷⁴ Dudenredaktion (ed.2014): op. cit., S. 22.

¹¹⁷⁵ Duden (ed.): *Wortfeld*. Online Publikation: <https://www.duden.de/rechtschreibung/Wortfeld> (15.11.2017).

¹¹⁷⁶ Vgl. Dudenredaktion (ed.2014): op. cit., S. 19.

dem Sprecher dadurch ermöglichen, seine Äußerungen zu variieren und damit Texte ansprechender zu gestalten. Im schulischen Kontext spielt die Unterscheidung in totale oder partielle Synonymie ebenso eine eher untergeordnete Rolle. Hier ist ausschlaggebend, dass Schüler verschiedene Begriffe zur Hand haben, die sie in ihren eigenen Sprachproduktionen gegeneinander austauschen können, um sich in ihren Texten seltener wiederholen zu müssen.

4.1.3.1. Ermittlung von Synonymen anhand des CdE

Die Suche nach Synonymen ist, wie bereits erwähnt, eine besondere Eigenschaft des CdE. Diese Suchoption wird fast ausschließlich in den Korpora von Mark Davies angeboten und führt zu vielen interessanten und brauchbaren Ergebnissen:

Syntax:	Bedeutung:	Beispiel:	Ergebnisse:
[=Lexem]	Synonyme des jeweiligen Lexems	[=ir]	partir, salir, presentar, andar
		[=mujer]	persona, joven, señora, esposa
[[=Lexem]]	alle morphologischen Formen der Synonyme des jeweiligen Lexems	[[=mujer]]	personas, joven, persona
Lexem1 [[=Lexem2]]	alle morphologischen Formen der Synonyme von Lexem2 als Teil einer Phrase	muy [[=guapo]]	muy bonito, muy bonita, muy hermosa, muy lindo
[Wortart*] [[=Lexem]]	ein Lexem der entsprechenden Wortart zusammen mit allen morphologischen Formen aller Synonyme des jeweiligen Lexems	[V*] [[=lentamente]]	caminando despacio, dijo pausadamente, vamos despacio
[=Lexem] SEC1= Jahrhundert SEC2= Jahrhundert	Vergleich der Häufigkeit der Synonyme des Lexems in zwei verschiedenen Jahrhunderten	[=guapo] SEC1= s20 SEC2= s13	SEC1: bonito, elegante, atractivo, genial SEC2: apuesto, valiente, hermoso
[=Lexem] SEC1= Genre SEC2= Genre	Vergleich der Häufigkeit der Synonyme des	[=simpático] SEC1= ORAL SEC2= ACAD	SEC1: majo, gracioso, amable, agradable

	Lexems in den verschiedenen Genres		SEC2: seductor, expansivo, llano, sugestivo
--	------------------------------------	--	---

Abbildung 99: CdE:Orig – Möglichkeiten der Suche nach Synonymen.¹¹⁷⁷

Es kann demnach im CdE nicht nur nach den Synonymen einzelner Lexeme gesucht werden, sondern es sind auch Suchen nach sämtlichen morphologischen Formen der Synonyme sowie nach Synonymen unter Vorgabe der sie begleitenden Lexeme möglich. Außerdem kann nach Ausdrücken gesucht werden, bei welchen die Wortart des vorausgehenden Lexems festgelegt ist, und es können die Synonyme eines Lexems in verschiedenen Jahrhunderten sowie in unterschiedlichen Genres verglichen werden. Des Weiteren kann mit zwei eckigen Klammern gesucht werden. Bei [[=mujer]] dienen die inneren Klammern in Verbindung mit dem Gleichheitszeichen dazu, nach Synonymen des Substantivs *mujer* zu suchen. Die äußeren Klammern ermöglichen die Suche nach allen Formen der Synonyme. Der aktuelle Teil des CdE (siehe Kapitel 2.1.1.3.) ist bislang das einzige Referenzkorpus, in dem nach synonymen Ausdrücken gesucht werden kann.

Für Schüler ist die Suche nach sinnverwandten Lexemen im Korpus eine interessante Möglichkeit, reale Sprachdaten selbst zu untersuchen. Wenn man beispielsweise nach Synonymen für *guapo* sucht, so werden in Zusammenhang mit *gato* die Adjektive *hermoso*, *estupendo* und *bonito* gebraucht. Zusammen mit *chico* werden *estupendo*, *valiente*, *majo*, *hermoso*, *galán* und *bonito* verwendet und in Kombination mit *hombre* erscheinen die Adjektive *valiente*, *hermoso*, *atractivo*, *valeroso*, *bello*, *elegante* und *gallardo* an oberster Stelle. Diese Option stellt somit eine Kombination der Suche nach Synonymen und Kollokationen dar, weil nach passenden Synonymen für bestimmte Lexeme gesucht wird, also nach adjektivischen Kollokatoren mit der Bedeutung *guapo* für eine bestimmte substantivische Basis.

Die Möglichkeit, semantisch-orientierte Suchanfragen durchzuführen, ist eine der nützlichsten Optionen des CdE, welche dieses gleichzeitig von anderen Korpora abhebt. Man kann sich allerdings nicht nur die Häufigkeit und Verteilung aller Synonyme eines bestimmten Lexems anzeigen lassen, sondern eben auch alle Einheiten, die zusammen

¹¹⁷⁷ Vgl. Davies, Mark: *Corpus del Español. Ayuda: Sinónimos*. Online Publikation: <http://www.corpusdelespanol.org/> (22.05.2011).

4. Nutzung von Korpora in der fremdsprachlichen Unterrichtspraxis

mit diesen Synonymen auftreten, beispielsweise Substantive, die mit den Synonymen von *sagrado* verwendet werden, oder Adjektive mit den Synonymen von *imagen*.¹¹⁷⁸

Ein Beispiel hierfür ist der Vergleich zweier partiell synonymmer Lexeme. Stellt man im CdE:Orig (Version 2008) unter *Mostrar* die Option *Comparar* ein und vergleicht jegliche Substantive, die in der Umgebung von *guapo* und *bonito* vorkommen, so kommt man zu dem Ergebnis, dass die relevantesten Kombinationen mit *guapo* zusammen mit *mozo/-a*, *chico/-a(s)*, *hombre*, *joven*, *muchacha* und *mujer(es)* gebildet werden, also ausschließlich mit Substantiven, die sich auf Personen beziehen, während *bonito* zwar auch mit *muchacha* und *niña*, ansonsten aber mit Dingen kombiniert wird (*cosa(s)*, *oliva*, *ojos*, *nombre*, *palabras*, *caras* und *voz*).

WORD 1 (W1): GUAPO (0.30)					WORD 2 (W2): BONITO (3.35)						
	WORD	W1	W2	W1/W2	SCORE		WORD	W2	W1	W2/W1	SCORE
1	MOZA	12	0	24.0	80.5	1	COSAS	26	0	52.0	15.5
2	MOZO	16	1	16.0	53.7	2	MUCHACHAS	14	0	28.0	8.3
3	CHICAS	7	0	14.0	47.0	3	COSA	13	0	26.0	7.7
4	CHICA	7	0	14.0	47.0	4	NIÑA	11	0	22.0	6.6
5	CHICO	4	0	8.0	26.8	5	OLIVA	9	0	18.0	5.4
6	HOMBRE	4	2	2.0	6.7	6	OJOS	7	0	14.0	4.2
7	JOVEN	4	3	1.3	4.5	7	NOMBRE	6	0	12.0	3.6
8	MUCHACHA	5	6	0.8	2.8	8	PALABRAS	6	0	12.0	3.6
9	MUJER	5	15	0.3	1.1	9	CARAS	5	0	10.0	3.0
10	MUJERES	4	13	0.3	1.0	10	VOZ	5	0	10.0	3.0

Abbildung 100: CdE:Orig (Version 2008) – Suche nach Substantiven, die häufig mit *guapo* oder *bonito* kombiniert werden.¹¹⁷⁹

Die häufig gehörte Frage «¿Qué es la diferencia entre ... y ...?» kann auf diese Weise auch einmal den Schülern gestellt werden. Diese können somit die Tatsache, dass *guapo* mit Personen und *bonito* vor allem mit Sachen verwendet wird, selbst herausfinden. Auf diese Weise könnten die Schüler auch *esquina* und *rincón* oder das in Kapitel 2.1.1.3. erwähnte Paar *empezar* und *comenzar* einander gegenüberstellen. Die Schüler können derartige Suchen auch zu Hause durchführen, sofern die Klasse bereits in die Funktionsweise des Korpus eingeführt wurde.

4.1.3.2. Erstellung eines Wortfelds anhand des CdE

Die Option der Suche nach Synonymen im CdE eröffnet neue Möglichkeiten für den Spanischunterricht, denn es ist somit möglich, Ausdrücke mit ähnlicher Bedeutung zu suchen, die zusätzlich zu Kollokationen in eine Mindmap eingetragen und für eine Personenbeschreibung genutzt werden können. Ein Beispiel hierfür ist die Suche nach unterschiedlichen Ausdrücken zum Thema ‚Bart‘. Hier können mit dem Suchausdruck

¹¹⁷⁸ Vgl. Davies: *Corpus del Español – Five Minute Tour*. Op. cit. (20.05.2011).

¹¹⁷⁹ Davies: *Corpus del Español – Género/Histórico (Version 2008)*. Op. cit. (09.11.2017).

[=barba] folgende Ergebnisse ermittelt und daraufhin in eine umfangreiche Mindmap zum Thema ‚Gesicht‘ integriert werden:

CONTEXTO	FREC
BARBA [S]	1074
BARBILLA [S]	135
PERA [S]	108
PERILLA [S]	58
CHIVA [S]	18
SOTABARBA [S]	3
TOTAL	1396

Abbildung 101: CdE:Orig – Suche nach Synonymen von *barba*.¹¹⁸⁰

Ein weiteres Beispiel ist die Behandlung der *verba dicendi*. Hier wird von den Schülern erwartet, bei der Übertragung von direkter zu indirekter Rede nicht nur die korrekten Formen des Verbs *decir*, sondern unterschiedliche Lexeme als Einleitungsverben für die Sätze in der indirekten Rede zu verwenden. Als Vorbereitung darauf ist es daher sinnvoll, die Schüler im CdE:Orig nach Synonymen von *decir* suchen zu lassen. Daraufhin sollen sie mindestens zehn der angezeigten Verben in ihrem Text in der indirekten Rede verwenden.

CONTEXTO	FREC
DECIR [S]	23937
HABLAR [S]	8347
CONTAR [S]	2617
EXPLICAR [S]	1529
OBSERVAR [S]	1289
ASEGURAR [S]	1226
SOSTENER [S]	1123
EXPRESAR [S]	832
AFIRMAR [S]	781
SEÑALAR [S]	765

Abbildung 102: CdE:Orig – Suche nach Synonymen von *decir*.¹¹⁸¹

Anhand dieser Ergebnisse kann zudem ein Wortfeld wie das folgende erstellt werden:

¹¹⁸⁰ Davies: *Corpus del Español – Género/Histórico*. Op. cit. (03.11.2017).

¹¹⁸¹ Davies: *Corpus del Español – Género/Histórico*. Op. cit. (03.11.2017).

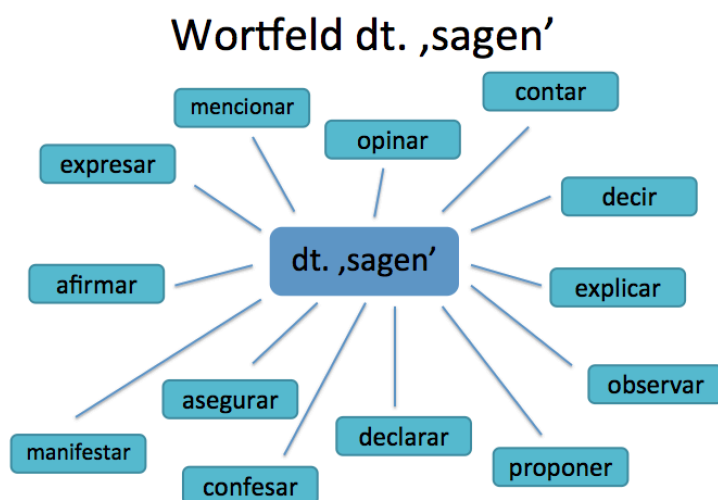


Abbildung 103: Spanisches Wortfeld zum deutschen Begriff ‚sagen‘.

Wenn es den Schülern nun bei der Bearbeitung von Übungen zur indirekten Rede an einem Einleitungsverb fehlt, können sie das Wortfeld zurate ziehen. Wortfelder zu unterschiedlichsten Themen können anhand des CdE auf diese Weise relativ einfach erstellt werden.

4.1.3.3. Nutzung des CdE als Thesaurus

Um die Vielfalt der Ausdrucksmöglichkeiten der Schüler zu erhöhen, sind Suchanfragen nach Synonymen für Lexeme, die von Schülern relativ häufig verwendet werden, lohnenswert. Beispielsweise kann nach Synonymen für das Substantiv *problema*, das von Schülern sehr häufig in ihre Aufsätzen eingebaut wird, im CdE:Orig gesucht und ein passendes sinnverwandtes Lexem für den jeweiligen sprachlichen Kontext ausgewählt werden.

CONTEXTO	FREC
DUDA [S]	8028
PREGUNTA [S]	7854
PROBLEMA [S]	6501
ASUNTO [S]	4495
CUESTIÓN [S]	4384
ELECCIÓN [S]	2137
EJERCICIO [S]	2116
CONFLICTO [S]	2026
DIFICULTAD [S]	1749
COMPROMISO [S]	1345

Abbildung 104: CdE:Orig – Suche nach Synonymen von *problema*.¹¹⁸²

¹¹⁸² Davies: *Corpus del Español – Género/Histórico*. Op. cit. (03.11.2017).

Ebenso können synonyme Ausdrücke für Verben anhand des CdE:Orig gesucht werden. So ergeben sich für das Verb *preguntar* beispielsweise *buscar*, *averiguar*, *dudar*, *investigar*, *consultar*, *rogar*, *interrogar* und *cuestionar*. Die Schüler sollten derartige Suchen am besten direkt vor oder während des Verfassens eines Aufsatzes durchführen und sie sollten dazu angehalten werden, in ihren Aufsätzen mehr zu variieren sowie Wiederholungen zu vermeiden.

4.1.3.4. Verbesserung der Ausdrucksfähigkeit anhand von Synonymen

Zudem haben Lerner des Öfteren Schwierigkeiten damit, in Textproduktionen unterschiedliche Ausdrücke zu verwenden. So nutzen sie einige Adjektive wie *grande*, *importante*, *alto* und *bueno* sehr häufig und variieren selten. Anhand des Arbeitsblatts 6 – *Sinónimos para adjetivos corrientes* (Anhang 15) sollen die Schüler daher andere Adjektive für vorgegebene Kombinationen aus einem Substantiv und einem frequenten, „gewöhnlichen“ Adjektiv finden. Dieses Arbeitsblatt ist eine Präzisierung des Arbeitsblatts 4 – *Adjetivos adecuados* (siehe Kapitel 4.1.1.3.4., Anhang 13). Während bei AB 4 generell nach frequenten Kombinationen aus Substantiven und Adjektiven zur Verbesserung der Wortschatzkenntnisse gesucht wird, sind es bei AB 6 synonyme Lexeme für bestimmte Adjektive, die besonders häufig verwendet werden. Es handelt sich bei beiden Arbeitsblättern um annähernd dieselbe Gruppe von Substantiven und um eine ähnliche Vorgehensweise. Nutzen Schüler gehäuft die üblichen Adjektive *grande*, *importante*, *alto* und *bueno*, so ist AB 6 besser geeignet, sofern die Schüler generell die Vielfalt sprachlicher Mittel ausbauen sollen oder im Allgemeinen selten Adjektive verwenden, ist AB 4 zu empfehlen. Die Arbeitsblätter können hinsichtlich der Auswahl der Substantive und Adjektive an die Bedürfnisse der Klasse angepasst werden.

Um passende Adjektive zu ermitteln, soll der Schüler das jeweils vorgegebene Substantiv im CdE:Orig (Version 2008) in eckigen Klammern eingeben, beispielsweise [ventaja], um nach Treffern im Singular und Plural zu suchen.

4. Nutzung von Korpora in der fremdsprachlichen Unterrichtspraxis

Abbildung 105: CdE:Orig (Version 2008) – Suche nach Synonymen von *grande*, die zusammen mit dem Substantiv *ventaja* verwendet werden.¹¹⁸³

Zudem wird angegeben, dass im Umkreis von einem Lexem links und rechts des Suchbegriffs alle Synonyme des jeweiligen Adjektivs, im ersten Beispiel mittels des Ausdrucks [=grande], ermittelt werden. Bei der Suche im CdE:Orig (Version 2008) wird bei der Grundeinstellung automatisch im Bereich des 19. und 20. Jahrhunderts gesucht, also im Subkorpus des modernen Spanisch (siehe Kapitel 2.1.1.3.), daher ist hier keine Abänderung der Einstellungen vorzunehmen. Im Bereich *Ordenar y limitar* sollte bei *Ordenar* die Option *Relevancia* und bei *Mínimo* die Zahl drei als Mindestanzahl der gefundenen Treffer ausgewählt werden. Mithilfe dieses vierten Unterbereichs ist es möglich, die Ergebnisse nach Frequenz, Relevanz sowie alphabetisch zu ordnen. Vor allem bei Suchen, die sich auf das gemeinsame Auftreten zweier Lexeme beziehen, besteht ein großer Unterschied zwischen der Anordnung nach Frequenz und derjenigen nach Relevanz. Während unter den ersten Treffern bei der Anordnung nach Frequenz zwar sehr häufige, aber wenig kollokative Verbindungen angezeigt werden, erscheinen beim Kriterium der Relevanz signifikante Verbindungen zwischen Lexemen, welche als Kollokationen bezeichnet werden können (zum Begriff der Kollokation siehe Kapitel 4.1.1.). Gibt man beispielsweise den Suchausdruck * *frío* ein, so erscheinen bei der Anordnung nach Frequenz unter den ersten Ergebnissen *el frío*, *de frío*, *y frío*, *del frío* und *un frío*. Lässt man die Ergebnisse nach Relevanz anordnen, so sind die ersten fünf

¹¹⁸³ Davies: *Corpus del Español – Género/Histórico* (Version 2008). Op. cit. (09.11.2017).

Ergebnisse *sudor frío, sudando frío, clima frío, intenso frío* und *sepulcro frío*. Die Ergebnisse der Anordnung nach Frequenz sind leicht nachzuvollziehen, verdeutlicht man sich die häufige Verwendung von Funktionswörtern. Für Lernende des Spanischen sind Kollokationen, wie sie die Anordnung nach Relevanz zeigt, brauchbarer, da diese ihnen wichtige Wortverbindungen für ihre Sprachkenntnisse liefern.

Sofern die angegebenen Einstellungen vorgenommen wurden, können die Schüler mit der Suche beginnen. Sie sollen jedoch nicht nur passende Adjektive für das angegebene Substantiv finden (beispielsweise synonyme Adjektive in der Bedeutung *grande* für das Substantiv *ventaja*), sondern zudem einen Beispielsatz angeben, damit sie auch lernen, wie die Adjektive kontextualisiert werden. Für die Formulierung der Beispielsätze können sie sowohl die Konkordanzen des Korpus als auch ein Online-Wörterbuch zu Hilfe nehmen. Diese Übung ist ab Anfang des dritten Lernjahrs geeignet und dient vor allem der Verbesserung der Ausdrucksfähigkeit, um den Schülern zu mehr Abwechslung in ihren Textproduktionen zu verhelfen.

Die Suche nach Synonymen anhand von Korpora erscheint im Spanischen deshalb so vielversprechend, weil es für Lerner dieser Sprache keine Selbstverständlichkeit ist, im Rahmen von Textverarbeitungsprogrammen ein Programm wie *Thesaurus* für die englische Sprache an der Hand zu haben, das beim Verfassen von Texten Synonyme anbietet. Das CdE ermöglicht es jedoch, binnen weniger Sekunden eben solche Ausdrücke zu ermitteln, um die eigene Ausdrucksfähigkeit in der Fremdsprache zu verbessern. Ob die vorgeschlagenen Ausdrücke tatsächlich in den jeweiligen sprachlichen Kontext passen, muss der Lerner selbst abwägen, doch ein Vorteil des Korpus gegenüber dem *Thesaurus* ist, dass die Konkordanzen der synonymen Ausdrücke abgerufen werden können und somit der passende Ausdruck ausgewählt werden kann. Im *Thesaurus* muss der Nutzer ohne weitere Anhaltspunkte entscheiden, ob das vorgeschlagene Lexem in den jeweiligen sprachlichen Kontext passt, was für Lerner einer Fremdsprache durchaus ein Problem darstellen kann. Außerdem ist es möglich, die Kollokatoren zweier synonymen Lexeme zu vergleichen, um Unterschiede in deren Bedeutung und Gebrauch zu ermitteln, beispielsweise zwischen *poderoso* und *potente*. Dieser Vergleich von Kollokatoren geht über die Arbeit eines Thesaurus hinaus und ermöglicht die Ermittlung selbst kleinster Unterschiede.

4.1.3.5. Klärung von Bedeutungs- und Verwendungsunterschieden

Anhand von Korpora können somit nicht nur sinnverwandte Lexeme ermittelt werden, sondern es ist auch möglich, Unterschiede zwischen bedeutungsähnlichen Einheiten herauszufinden. Mithilfe der Konkordanzen aus einem Korpus können die Schüler selbstständig eben diese unterschiedlichen Verwendungskontexte untersuchen und daraus Schlüsse für ihre eigene Sprachproduktion ziehen. Sie können selbst ermitteln, in welchen Kontexten das eine und in welchen das andere Lexem verwendet wird. Des Weiteren kann anhand des Korpus gesehen werden, welcher der beiden Ausdrücke häufiger vorkommt, und auch, in welchen thematischen Bereichen oder Textsorten sie verwendet werden.

Johns (1991) ließ Lerner Konkordanzen mit den Synonymen *convince* und *persuade* analysieren und ihre eigenen Schlussfolgerungen daraus ziehen. Es wurde schnell herausgefunden, dass auf *convince* in zehn der elf Konkordanzen ein *that-clause* und auf *persuade* in 14 der 18 Beispiele der *to-infinitive* sowie viermal ein *that-clause* folgte. Daraufhin wurde die Frage gestellt, worin sich die beiden Anschlüsse bei *persuade* unterscheiden, und hierfür gab es verschiedene Erklärungsversuche. Die Lehrkraft erklärte diese Tatsache damit, dass der Infinitiv bevorzugt wird, sofern es keinen Wechsel des Subjekts gibt, wohingegen ein *that-clause* stehen muss, wenn ein solcher Wechsel erfolgt. Ein Lerner schloss aufgrund der Konkordanzen allerdings darauf, dass der *to-infinitive* für gewöhnlich verwendet wird, wenn jemand davon überzeugt werden soll, etwas zu tun, und der *that-clause* in denjenigen Fällen steht, in denen jemand von etwas, also einer Tatsache, überzeugt werden soll. Somit können Lerner anhand von Konkordanzen durchaus eigene Leitsätze ableiten, doch Johns warnte in diesem Zusammenhang auch vor Übergeneralisierungen, wie bereits in Kapitel 1.1.1. erwähnt. Allein das Fehlen einer Struktur in einem Korpus heißt nicht, dass diese generell nicht existiert und daher ungrammatisch ist.¹¹⁸⁴

Das CdE kann folglich auch von den Schülern genutzt werden, um herauszufinden, wie sich Lexeme mit ähnlicher Bedeutung in ihrer Verwendung unterscheiden. So fällt es Lernern für gewöhnlich schwer, zwischen den Substantiven *posibilidad* und *oportunidad* zu unterscheiden. Obwohl die beiden Lexeme sich in ihrer deutschen Bedeutung (‚Möglichkeit‘ und ‚Gelegenheit‘) unterscheiden, werden sie dennoch häufig

¹¹⁸⁴ Vgl. Johns (1991): op. cit., S. 1–16.

verwechselt, da die Schüler umgangssprachlich meist den deutschen Begriff ‚Möglichkeit‘ für beide Lexeme verwenden, ohne auf die sprachliche Umgebung und die Bedeutung zu achten. Mittels der Funktion *Comparar* können die typischen Kontexte der Lexeme jedoch gegenübergestellt und verglichen werden. So ergibt die Suche nach den vier Einheiten links und rechts von *posibilidad* und *oportunidad* im CdE:Orig für Ersteres, dass es häufig zusammen mit den Verben *descartar* und *existir* gebraucht und unmittelbar neben *mis* oder *nuestros* genutzt wird, wie im Ausdruck *dentro de mis/nuestros posibilidades*. Der Begriff *oportunidad* fällt hingegen mit *tener* oder *aprovechar* sowie in Zusammenhang mit dem Ausdruck *igualdad de oportunidades*.

Auf diese Weise können auch die Synonyme *simple* und *fácil* miteinander verglichen werden. Während bei der Suche nach einem Lexem links und rechts des Suchbegriffs für *simple* vor allem die Substantive *vista (a simple vista)*, *mayoría* und *curiosidad* erscheinen, wird *fácil* häufig mit den Verben *ser* (meist im *Condicional* oder *Futuro simple*), *encontrar* und *hacer* sowie mit den Substantiven *tarea* und *acceso* verwendet. Anhand der Funktion *Comparar* des CdE:Orig kann somit auf einfache Weise herausgefunden werden, in welchem Kontext zwei Begriffe, die der Bedeutung nach ähnlich sind, gebraucht werden. Das Korpus liefert hierfür ausreichend Belege, um die jeweiligen Lexeme einander gegenüber stellen und miteinander vergleichen zu können.

4.1.4. Homonyme und Polyseme

Neben Synonymen sind auch Homonyme und Polyseme von besonderem Interesse, nicht nur in der Linguistik, sondern auch im Fremdsprachenunterricht. In Bußmanns *Lexikon der Sprachwissenschaft* wird bezüglich der Beziehung zwischen Polysemie und Homonymie Folgendes angegeben:

„Zur Unterscheidung von P. und Homonymie wird traditionell das Kriterium der etymologischen Verwandtschaft herangezogen: Homonyme Ausdrücke lassen sich auf verschiedene historische Wurzeln zurückführen, polyseme Ausdrücke haben hingegen eine gemeinsame Wurzel. Zumindest für die synchrone Sprachbetrachtung ist die etymologische Verwandtschaft jedoch weitgehend irrelevant und zudem teilweise auch ungeklärt.“¹¹⁸⁵

Die Themen der Polysemie und der Homonymie werden aus diesem Grund in der vorliegenden Arbeit wie bei Pöll gemeinsam behandelt, da sich die Abgrenzung der

¹¹⁸⁵ Bußmann (2008): op. cit., S. 538.

beiden Phänomene als relativ komplex gestaltet und auch die Homonymie, obwohl es sich hier um keine semantische Thematik handelt, zu Umstrukturierungen des Wortschatzes führen kann.¹¹⁸⁶

Der Homonymie schreibt Pöll die folgenden Paare zu: *papa* (< Quechua *pápa*) ‘Kartoffel’ / *papa* ‘Papst’ (< lat. PAPA), *junco* (< lat. IUNCU) ‘Binse’ / *junco* (< Malayo) ‘Dschunke’, *cola* ‘Schwanz’ (< lat. CAUDA) / *cola* ‘Klebstoff’ (< gr. kólla), *carpa* ‘Karpfen’ (< lat. CARPA) / *carpa* ‘Zelt’ (in Hispanoamerika) (< Quechua), *vela* ‘Kerze’ (zu *velare* < lat. VIGILARE) / *vela* ‘Segel’ (< lat. VELA).¹¹⁸⁷

Die Einordnung als Polysemie oder Homonymie ist in vielen Fällen aber für Muttersprachler widersprüchlich, wie bei der Zuordnung des Substantivs *pupila*, was ‘Mündel’ und ‘Pupille’ bedeuten kann, zur Polysemie. Ebenso verhält es sich mit den Homonymen *vago* ‘vacío, desocupado’ (< lat VACUU) und *vago* ‘indeciso, indeterminado’ (lat. < VAGU). Auch bei den beiden Bedeutungen von *sueño* ‘Schlaf’ und ‘Traum’ handelt es sich um Homonyme, wobei hier aber leicht angenommen werden kann, dass polyseme Lexeme vorliegen.¹¹⁸⁸ Laut Blank handelt es sich bei den beiden Bedeutungen von *sueño* um sekundäre Polysemie, da die ursprüngliche Homonymie in eine metonymische Polysemie uminterpretiert wurde.¹¹⁸⁹

Ob im jeweiligen Fall polyseme oder homonyme Lexeme vorliegen, kann im fremdsprachlichen Unterricht oft nicht geklärt werden. Zudem spielt es meist keine Rolle, ob die beiden Lexeme eine etymologische Verwandtschaft verbindet. Somit werden die Paare von Lexemen im folgenden Kapitel als Homonyme bezeichnet, um damit nicht den Anschein zu erwecken, die Lemmata entsprängen einer gemeinsamen Wurzel.

Eine Besonderheit der Homonyme ist die Tatsache, dass sich viele Lexeme nur durch deren Geschlecht unterscheiden, wie im obigen Fall von *papa*. Daher erkennt man die jeweilige Bedeutung bei diesen Paaren abgesehen vom sprachlichen Kontext auch an Artikeln oder Adjektiven, die zusammen mit dem Substantiv gebraucht werden. Ein weiteres polysemes Paar von Lexemen, welches zu Verwechslungen führen kann, ist *la cabeza* (,der Kopf‘) und *el cabeza* (,das Oberhaupt‘, beispielsweise das Familienoberhaupt).

¹¹⁸⁶ Vgl. Pöll (2002): op. cit., S. 76.

¹¹⁸⁷ Vgl. Pöll (2002): op. cit., S. 77.

¹¹⁸⁸ Vgl. Pöll (2002): op. cit., S. 77-78.

¹¹⁸⁹ Vgl. Blank (2001): op. cit., S. 112.

Um den Schülern den Unterschied zwischen den verschiedenen Bedeutungen von *cabeza* nahezubringen, könnte die Lehrkraft die Schüler im CdE:Nuevo (siehe Kapitel 2.6.1.) forschen lassen oder ihnen eine Sammlung von Konkordanzen vorlegen und sie selbst die möglichen Bedeutungen dieses Substantivs anhand des sprachlichen Kontexts ermitteln lassen. Für eine Suche nach *cabeza* mit unterschiedlichen Artikeln wird das Substantiv im Reiter *Colocados* als *Palabra/frase* eingegeben, im Feld *Colocados* die Option *art.TODOS* sowie ‚ein Lexem links des Suchbegriffs‘ als Intervall ausgewählt.

Abbildung 106: CdE:Nuevo – Suche nach *cabeza* mit unterschiedlichen Artikeln.¹¹⁹⁰

Im Folgenden wird eine Liste mit den verschiedenen Artikeln angezeigt, die mit *cabeza* erscheinen, sowie deren jeweilige Frequenz. Im CdE:Nuevo werden über 175.000 Treffer für die feminine Form *la cabeza* aufgelistet und lediglich 887 für *el cabeza*.

mí, que no puedes, que te duele la cabeza , que no tienes ganas... # # # Llega una señora
jefe, hic, ¿cuantos chichones tengo en la cabeza ? - Doce, ¿por quÚ? - ¡Ah!, hic,
en la carpintería. Por un descuido Pedro mete la cabeza en la sierra y se corta una oreja. - ¡Ayyyyy mi oreja,
el partido de el miércoles, Riquelme le tocó la cabeza a Orion. Román no dijo, mostró. Se trataba de la redención
nada mas. La gente de Independiente debe levantar la cabeza y dejar de protestar, a mi como hincha de Central, me doliò
quien tuvo que ser hospitalizado por un corte en la cabeza . Son amargos si, antes eran las viudas, ahora somos los bosteros
De Pablo, su asesor, le están llenando la cabeza en contra de mí. Estos tipos, cuando yo todavía estaba al lado

Abbildung 107: CdE:Nuevo – Konkordanzen des Substantivs *cabeza* mit definitem, femininem Artikel.¹¹⁹¹

Anhand der Konkordanzen ist klar ersichtlich, welche Bedeutung für das jeweilige Lexem mit dem femininen und dem maskulinen Artikel zutrifft.

derecho era tuitivo respecto a ella: si faltaba el cabeza de familia, que mandaba, tenía todo a su nombre, ganaba el
formación, o cuando el nivel de estudios de el cabeza de familia es menor. El visionado en familia, independientemente de cuál sea
bastante fuleros como la A con los ojitos o el cabeza de picos. (Algunos excelentes y otros desastrosos, ¿cuál te gusta
lectores tengan un poco mas de sentido común que el cabeza de chorlito que ha perdido su tiempo en este seudo artículo. Mi primer
vivían abuelos, tíos, primos... El cabeza de familia se hacía cargo de todos, y a su manera, cada
. Pienso lo mismo y creo que esto para el cabeza a cabeza que viene en las elecciones le puede dar la victoria. recien
. he sentido su presencia, he visto como el cabeza de grupo palidecia a diario, y dijo escuchar voces, hasta q en
el mayor de los hermanos y aunque debería ser el cabeza de familia y asegurar la protección de sus hermanas, la muerte de su

Abbildung 108: CdE:Nuevo – Konkordanzen des Substantivs *cabeza* mit definitem, maskulinem Artikel.¹¹⁹²

¹¹⁹⁰ Davies: *Corpus del Español – Web/Dialectos*. Op. cit. (16.12.2016).

¹¹⁹¹ Davies: *Corpus del Español – Web/Dialectos*. Op. cit. (16.12.2016).

¹¹⁹² Davies: *Corpus del Español – Web/Dialectos*. Op. cit. (16.12.2016).

Die Schüler lernen im Unterricht zunächst nur die Form *la cabeza* kennen, da das maskuline Pendant für schulische Zwecke relativ unbedeutend ist. Anhand dieses Beispiels wird für die Schüler allerdings auch ersichtlich, welches der beiden Lexeme für die Kommunikation relevanter ist. Zudem kann mittels derartiger Forschungen im Korpus auf mögliche Fehlerquellen hingewiesen werden, da im Fall der Verwendung des anderen Genus nicht nur der falsche Artikel ausgewählt wurde, sondern sich eine ganz andere Bedeutung ergibt.

Eine umfassende Verwendungsmöglichkeit der Korpora auf diesem Gebiet bietet Arbeitsblatt 7 – *Homónimos* (Anhang 16). Anhand des CdE:Nuevo sollen die Schüler durch die Suche der Substantive in Kombination mit einem Artikel herausfinden, mit welchem Genus die Lexeme die angegebene Bedeutung haben. Hierfür geben die Schüler das jeweilige Lexem im Reiter *Colocados* in das Feld *Palabra/frase* ein, wählen *art.Todos* im Bereich *Colocados* aus und bestimmen ‚ein Lexem links des Suchbegriffs‘ als Intervall. Bei dieser Suche werden alle Artikel der Frequenz nach aufgelistet, der Schüler wählt einen maskulinen oder femininen Artikel aus, lässt sich die Konkordanzen anzeigen und bestimmt anhand des Kontexts, ob es sich hierbei um die erste oder zweite Bedeutung handelt. Nun kann er das Feld der jeweiligen Bedeutung in der passenden Farbe markieren. Danach wählt er den jeweils anderen Artikel aus, überprüft, ob die andere Bedeutung auf dessen Konkordanzen zutrifft, und markiert das andere Feld. Zum Schluss wird das Feld *más usado* in derjenigen Farbe markiert, welche das Genus mit dem häufigsten Vorkommen im Korpus darstellt.

Die farbliche Markierung der Felder lockert die Aufgabe zum einen auf, zum anderen können sich die Schüler die farblich unterlegten Bedeutungen besser einprägen. Die Wahl der Farben ist hier zudem zu beachten, da rot/rosa gewöhnlich für weibliche und dunkel- oder hellblau für männliche Substantive steht, was den Schülern aufgrund ihres Weltwissens zusätzlich bei der Verankerung der jeweiligen Bedeutung hilft. Dieses Arbeitsblatt ist ab Ende des zweiten Lernjahrs geeignet. Zur Suche nach dem jeweiligen Genus bei Homonymen ist zu sagen, dass diese Aufgabe den Schülern sehr leicht fällt. Sofern die Sucheinstellungen einmal eingegeben und ein oder zwei gemeinsame Suchen durchgeführt wurden, können die Schüler vollkommen selbstständig arbeiten und benötigen keinerlei Hilfestellung mehr. Aus diesem Grund ist die Bearbeitung von

Arbeitsblatt 7 im Computerraum eine willkommene Abwechslung im Unterrichtsalltag und wird von den Schülern auch als solche aufgefasst.

Die Schüler könnten dieses Arbeitsblatt auch mithilfe eines Wörterbuchs ausfüllen, doch die Beschäftigung mit den Konkordanzen des Korpus gibt den Schülern einerseits einen Überblick darüber, in welchem sprachlichen Kontext das jeweilige Lexem verwendet wird, und zum anderen einen Anhaltspunkt, wie häufig die Formen im alltäglichen Spanisch vorkommen. So ist beispielsweise sofort ersichtlich, dass *el cabeza* eine relativ seltene Form ist, wohingegen *la cabeza* sehr häufig gebraucht wird. In Wörterbüchern sind zwar die einzelnen Bedeutungen der Lexeme aufgelistet, es finden sich allerdings keinerlei Informationen zur Frequenz. Unter den angegebenen Paaren sind in der Regel ein recht häufig gebrauchtes und ein eher seltenes Lexem vertreten, sodass die Schüler normalerweise nur eines davon für ihre Sprachproduktionen benötigen. Aus diesem Grund sollen die Schüler auch angeben, welche Bedeutung häufiger und welches Genus somit frequenter ist. Die Übung weist hier auf mögliche Fehlerquellen hin, wie sie Spanischlehrer beispielsweise im Fall von *el / la capital* bei Schülertexten über lateinamerikanische Länder bereits zur Genüge gelesen und verbessert haben. Zusätzlich kann die Lehrkraft auch die genaue Anzahl der Vorkommnisse angeben lassen, wobei es zu beachten gilt, dass es sich beim CdE:Nuevo um ein Monitorkorpus handelt (siehe Kapitel 2.7.), das heißt, dass sich die genaue Anzahl ändern kann, wenn neue Texte hinzugefügt werden. Es kommt hier jedoch nicht auf die exakte Anzahl an, sondern darauf, dass die Schüler erkennen, in welcher Bedeutung das Lexem häufiger gebraucht wird. Die ermittelten Zahlen können zudem gerundet werden.

Dieses Arbeitsblatt könnte auch mit den Korpora CREA (siehe Kapitel 2.1.1.1.) oder CORPES (siehe Kapitel 2.1.1.2.) bearbeitet werden, jedoch muss hier die Suchanfrage auf eine andere Weise gestellt werden. In der Benutzeroberfläche wird das jeweilige Substantiv (*cabeza*) als Lemma eingegeben und die Option *Proximidad* geöffnet. Im Bereich *Proximidad* wird der Artikel als Form eingegeben und ‚ein Lexem links des Suchbegriffs‘ als Intervall bestimmt. Zusätzlich können die beiden Wortarten *sustantivo* und *artículo* im jeweiligen Unterbereich ausgewählt werden, um die Suche möglichst genau durchzuführen. Durch einen Klick auf *Concordancia* werden auch hier einige Konkordanzen angezeigt, anhand derer die Bedeutung ermittelt werden kann. Theoretisch würden die beiden Korpora der *Real Academia Española* auch die

Möglichkeit anbieten, unter Angabe des jeweiligen Genus nach dem Substantiv zu suchen und somit ohne die zusätzlichen Einstellungen im Bereich *Proximidad* zu arbeiten. Nach ein paar Beispielen wird hier allerdings klar, dass bei der automatischen morphosyntaktischen Annotation (siehe Kapitel 1.4.5.) immer wieder falsch kategorisiert wird, so auch bei der Einordnung nach Genus. Die Lehrkraft müsste die Klasse daher zuvor darauf hinweisen, zusätzlich genau auf die in den Beispielen erscheinenden Artikel zu achten, um ihre Zuordnung nicht aufgrund falsch zugewiesener Konkordanzen zu treffen. In diesem Fall ist die Suche mittels der Option *Proximidad*, bei der nach einem direkt vorausgehenden Artikel *el* oder *la* gesucht wird, genauer und zielführender, obgleich nur nach denjenigen Sätzen mit dem jeweiligen Artikel und nicht nach allen femininen oder maskulinen Formen generell gesucht werden kann. Ebenso kann mit der alten Version des CREA gearbeitet werden. Hier wird der Suchausdruck, beispielsweise *la cabeza*, insgesamt eingegeben. Allerdings ist die Anzeige im CREA (Version 2015) beschränkt, wodurch bei frequenten Lexemen häufig keine Konkordanzen angezeigt werden können.

Ein zusätzlicher Vorteil der Verwendung des CdE:Nuevo bei der Ermittlung der Bedeutung von Homonymen gegenüber den anderen Korpora ist die schnelle Bearbeitung. Im CdE:Nuevo wird zunächst nach allen mit dem Substantiv erscheinenden Artikeln gesucht und im Anschluss lässt sich der Nutzer Konkordanzen mit dem weiblichen und dem männlichen Artikel anzeigen. Auf diese Weise wird pro Substantiv nur eine Suche, allerdings mit zwei Anzeigeoptionen durchgeführt, wohingegen im CREA oder im CORPES zwei Suchen durchgeführt werden müssen. Zudem sind die Konkordanzen im CdE:Nuevo aus Internettextrten entnommen, wodurch sie relativ aktuell sind und nicht aus literarischen Texten stammen, welche die Schüler häufig nicht verstehen.

Es gilt jedoch bei der Verwendung des CdE:Nuevo Folgendes zu beachten: Speziell bei dieser Suche nach Substantiven mit Artikel werden auch Beispiele aufgelistet, in denen eine Kombination aus den Präpositionen *a* oder *de* und dem definiten maskulinen Artikel *el* vorkommt. Das CdE:Nuevo führt diese Formen in den farblich hervorgehobenen Konkordanzen des Suchbegriffs (*el cabeza* oder *la cabeza*) allerdings getrennt als *de el* und *a el* auf und nicht als *del* und *al*. Diese Tatsache sollte den Schülern vorab mitgeteilt werden, um Verwirrung zu vermeiden. Die Schüler können sich

andererseits auch die Konkordanzen mit dem unbestimmten Artikel anzeigen lassen.

Abgesehen von dieser umfassenden Aufgabe zu homonymen Paaren von Lexemen, kann der Lehrer die Schüler allerdings auch nach polysemen Lexemen, wie beispielsweise *estrella*, im Korpus suchen und deren Bedeutungen ermitteln lassen. Bereits bei den ersten vier Konkordanzen dieses Substantivs im CdE:Nuevo wird ersichtlich, dass dieses Lexem verschiedene Bedeutungen hat, wie Himmelskörper oder berühmte Persönlichkeit:

Maria, Córdoba. Juan Olcese Admirada Rosario: Que puedas atrapar esta Navidad la estrella de Belen, para que bendiga todos tus anhelos y proyectos
lo proyecte como crack? ¿ Por qué Independiente contrata un refuerzo sin impronta de estrella y le paga el doble que el sueldo promedio de un futbol
que cualquiera que viaja en avión debe notificar acerca de su profesión. Podría ser estrella de rock: a juzgar por sus movimientos antes y después de la
jueves 11 / 4 en un asado en Caseros, en un club barrial de estrella bien dorada, con el Loco Luis como anfitrión, y al que concurrieron

Abbildung 109: CdE:Nuevo – Konkordanzen für das Substantiv *estrella*.¹¹⁹³

Anhand derartiger Suchanfragen und Analysen von Konkordanzen in einem Korpus arbeiten die Schüler nicht nur mit realen Beispielen und forschen selbst im Korpus, um Bedeutungen herauszufinden, sondern sie erweitern damit ebenso ihre Lernstrategien. Sie erwerben Wissen bezüglich der Funktionsweise von Korpora und können diese auch für eigene Zwecke bei ihren Textproduktionen zu Hause nutzen.

Morphosyntaktisch annotierte und lemmatisierte Korpora sind jedoch nicht nur zur Ermittlung des Bedeutungsunterschieds zwischen polysemen und homonymen Lexemen von Nutzen, sondern sie lassen ebenfalls die Unterscheidung zwischen Lexemen zu, die lediglich in bestimmten sprachlichen Kontexten formgleich sind. Wenn man beispielsweise nach dem Lexem *trabajo* ohne Angabe der Wortart sucht, so zeigt das CdE sowohl Beispiele für das maskuline Substantiv *trabajo* im Singular als auch für das Verb *trabajar* in der 1. Person Singular Indikativ Präsens Aktiv. Sofern man nun die jeweilige Wortart auswählt, werden nur Beispiele für das Substantiv oder die Verbform angezeigt. Des Weiteren kann mithilfe der Auswahlmöglichkeiten im Feld der Wortart nach bestimmten Wortformen gesucht werden, auch wenn diese sich auf dieselbe Wortart beziehen. So können unterschiedliche Zeitformen der 1. Person Plural des Verbs *cantar* unterschieden werden. Das Verb *cantar* hat in der 1. Person Plural sowohl im *presente de indicativo* als auch im *pretérito indefinido de indicativo* die Form *cantamos*. Sofern bei der Wortart *vPRES* oder *vPret* eingestellt wird, werden nur die Formen einer der beiden Zeiten unter den Ergebnissen aufgezählt. Allerdings muss auch hier erwähnt werden, dass

¹¹⁹³ Davies: *Corpus del Español – Web/Dialectos*. Op. cit. (16.12.2016).

die automatische morphosyntaktische Annotation und Lemmatisierung derartige Lemmata teilweise falsch kategorisiert und daher einige Konkordanzen falsch zugeordnet werden.

Auch im CORPES kann zwischen derartigen formgleichen Lexemen unterschieden werden, wobei die logischen Operatoren Y (beide Kriterien), O (eines der beiden Kriterien) und NO (eines aber nicht das andere Kriterium) komplexe Eingaben erleichtern. Beispielsweise müsste es theoretisch möglich sein, mit *Lema ‚parar‘ NO Lema ‚para‘* unter Angabe der jeweiligen Wortart ausschließlich nach *para* als Verbform von *parar* zu suchen. In der Praxis stellt sich jedoch auch hier die automatische Annotation und Lemmatisierung als problematisch heraus aufgrund falscher Kategorisierungen.

4.1.5. Wortbildung

Im Bereich der Wortbildung können Korpora ebenso für den Unterricht genutzt werden. Einige Korpora, unter anderem das CdE (siehe Kapitel 2.1.1.3., 2.2.1.3. und 2.6.1.) und die Korpora der *Real Academia Española* (siehe Kapitel 2.1.1.1., 2.1.1.2., 2.2.1.1. und 2.2.1.2.), bieten Suchoptionen mit Platzhaltern. Bei der Suche mit Platzhaltern kann sowohl mit Asterisken als auch mit Fragezeichen gearbeitet werden, wie anhand der folgenden Tabelle zu erkennen ist:

Syntax:	Bedeutung:	Beispiel:	Ergebnisse:
x?xx	Platzhalter: ? = genau ein Zeichen	familia?	familias, familiar
*xx	Platzhalter: * = beliebige Anzahl an Zeichen	*idad	realidad, autoridad
xxx		*segur*	seguro, seguridad, asegurar
Lexem*.[Wortart]	Platzhalter mit genauer Angabe der Wortart	*miento.[nn*]	movimiento, pensamiento

Abbildung 110: CdE:Orig – Möglichkeiten der Suche mit Platzhaltern.¹¹⁹⁴

Des Weiteren kann auch kombiniert mit Platzhaltern und nach Wortart gesucht werden, wie die letzte Option zeigt. Sofern alle Elemente des Suchausdrucks ohne Leerstelle geschrieben sind, wie beim obigen Beispiel **miento.[nn*]*, wird nach einem

¹¹⁹⁴ Vgl. Davies, Mark: *Corpus del Español. Sintaxis: Palabra/Frase*. Online Publikation: http://www.corpusdelespanol.org/help/syntax_s.asp (22.05.2011).

Lexem gesucht. Die Eigenschaften ‚endet auf *–miento*‘ und ‚Substantiv‘ sollen also auf dasselbe Lexem zutreffen. Wenn allerdings eine Leerstelle eingefügt wird, so wird nach einem Lexem gesucht, das auf *–miento* endet, und nach einem anderen, bei welchem es sich um ein Substantiv handelt. Sofern der Nutzer des CdE allerdings mit den Optionen der Eingabemaske arbeitet, wird der Suchausdruck vom System selbst generiert und es muss nicht derart genau auf die Zeichenfolge geachtet werden.

4.1.5.1. Präfigierung und Suffigierung

Die eben beschriebenen Suchoptionen können dazu genutzt werden, um den Schülern die verschiedenen Arten der Wortbildung nahezubringen. Eine Möglichkeit hierfür zeigt Arbeitsblatt 8 – *Formación de palabras* (Anhang 17), welches anhand des Korpus CREA bearbeitet wird (siehe Kapitel 2.1.1.1.). Um die gesuchten Lexeme zu finden, wird der angegebene Begriff (beispielsweise *justo*) zusammen mit den Platzhaltern ‚?*‘ in das Feld *Lema* eingegeben und die gesuchte Wortart (in diesem Fall *adjetivo*) ausgewählt. Da der Platzhalter ‚?‘ für ‚genau ein Zeichen‘ und der Platzhalter ‚*‘ für ‚kein, ein oder mehr als ein Zeichen‘ steht, kann mittels dieser Platzhalterkombination nach einem Prä- oder Suffix aus einem oder mehreren Zeichen gesucht werden. Als Ergebnis werden einige Beispiele mit dem Adjektiv *injusto* angezeigt, welches die Schüler als Lösung eintragen können. Während in der ersten Aufgabe nur nach Adjektiven gesucht wird, sollen die Schüler in der zweiten Aufgabe nach Adjektiven und Substantiven suchen. Um das passende Suffix zu ermitteln, werden die Platzhalter ‚?*‘ nun an den Stamm des angegebenen Lexems angehängt (beispielsweise *energ?**). Dieses Arbeitsblatt ist ab Ende des zweiten Lernjahrs geeignet und kann von den Schülern generell ohne Probleme bearbeitet werden. Nachdem die Lehrkraft den Schülern die Aufgabenstellung erklärt und anhand eines Beispiels veranschaulicht hat, können die Schüler selbstständig nach den passenden Prä- oder Suffixen suchen.

Die Lösungen zu den Arbeitsblättern stellen lediglich einen Lösungsvorschlag dar. Aufgrund der Tatsache, dass die meisten Korpora immer wieder erneuert oder erweitert werden, befinden sich auch jeweils andere Ergebnisse auf den ersten Seiten der Suchergebnisse. Da sich die Schüler vor allem die ersten Seiten anzeigen lassen und ihre Lösungen daraus ableiten, könnten manche Treffer noch nicht auf diesen Seiten angegeben sein. So befand sich beispielsweise in der zweiten Aufgabe kein Beispiel für

das Lexem *sentimiento* als Substantiv zum Verb *sentir* unter den ersten Treffern. Dies könnte sich jedoch ändern, sofern neue Texte in das Korpus integriert und andere Konkordanzen abgerufen werden. Die Ergebnisse sind also auch davon abhängig, welche Konkordanzen den Schülern angezeigt werden.

Arbeitsblatt 8 wird nicht nur mit dem CREA bearbeitet, sondern es ist auch mit diesem Korpus erstellt worden. Mithilfe des Suchausdrucks *?*ción* und der Suche nach *sustantivos* werden beispielsweise mit dem Suffix *-ción* gebildete Substantive ermittelt. Sofern die Lehrkraft noch mehr oder aber andere Lexeme in das Arbeitsblatt integrieren möchte, kann sie diese selbst im Korpus ermitteln. Das Arbeitsblatt kann ebenso mit dem CdE:Orig bearbeitet werden (siehe Kapitel 2.1.1.3.). Auch in diesem Korpus kann mit der angegebenen Platzhalterkombination gearbeitet und die gesuchte Wortart angegeben werden. Die jeweiligen Suchausdrücke (beispielsweise *?*justo*) können somit auf die gleiche Weise im CdE:Orig angewandt werden.

4.1.5.2. Wortfamilien

Generell kann in Korpora sehr gut nach Lexemen gesucht werden, welche einer Wortfamilie angehören, da der Einsatz von Platzhaltern die Suche nach Lexemen mit dem gleichen Wortstamm ermöglicht. Wie bereits in Kapitel 4.1.5. erklärt wurde, kann mithilfe des Platzhalters *,?’* nach Lexemen gesucht werden, die sich nur in einem Buchstaben unterscheiden. Ein Beispiel hierfür ist der Suchausdruck *familia?’*, welcher die Ergebnisse *familias* oder *familiar* liefert. Für die Suche nach Mitgliedern von Wortfamilien ist der Platzhalter *,*’* besser geeignet. So liefert die Suche nach *familia*’* die Ergebnisse *familia*, *familias*, *familiar*, *familiares*, *familiaridad*, *familiarmente*, *familiarizado*, *familiarizarse*, *familiarizados*, *familiariza* und viele weitere.

CONTEXTO	TOT
FAMILIA	10492
FAMILIAS	2436
FAMILIAR	1414
FAMILIARES	1153
FAMILIARIDAD	180
FAMILIARMENTE	86
FAMILIARIZADO	63
FAMILIARIZARSE	28
FAMILIARIZADOS	25
FAMILIARIZA	16
FAMILIARIZÁNDOSE	13
FAMILIARIZAR	11

Abbildung 111: CdE:Orig – Auflistung der Ergebnisse zum Suchausdruck *familia**.¹¹⁹⁵

So kann der Lehrer die Schüler mithilfe eines geeigneten Korpus, wie des CdE:Orig, des CREA oder des CORPES (siehe Kapitel 2.1.1.2.), selbstständig nach verwandten Lexemen suchen lassen, um beispielsweise neues Vokabular zu sammeln und einzuüben oder um sich wiederholende Strukturen in Aufsätzen durch ähnliche Ausdrücke zu ersetzen.

Eine Möglichkeit der Nutzung von Korpora für die Beschäftigung mit Wortfamilien wird anhand von Arbeitsblatt 9 – *Familias de palabras* (Anhang 18) veranschaulicht. Auf diesem Arbeitsblatt, das ab dem zweiten Lernjahr genutzt werden kann, sind einige Lexeme angegeben, welche unterschiedliche Formen der Wortbildung veranschaulichen, ohne jedoch thematisch in einer besonderen Beziehung zueinander zu stehen. Bei der Bearbeitung dieser Übung, sollen die Schüler anhand des CORPES ein Raster mit Wortfamilien vervollständigen. Dazu wird, wie auf dem Arbeitsblatt beschrieben, der Wortstamm zwischen Asterisken (beispielsweise **segur** für das Adjektiv *seguro*) als Lemma in die Suchmaske eingegeben und im Bereich *Clase de palabra* die jeweilige Wortart ausgewählt (beispielsweise *verbo*). Im Folgenden wird nun eine Liste mit Beispielen des Verbs *asegurar* angezeigt. Danach wird die nächste Wortart ausgesucht (beispielsweise *sustantivo*), woraufhin Beispielsätze für das Substantiv *seguridad* in der Liste der Ergebnisse erscheinen. Die Platzhalterkombination *,?** aus dem Kapitel 4.1.5.1. führt hier nicht zum Erfolg, da zuvor noch nicht bekannt ist, ob überhaupt ein Prä- oder Suffix angehängt wird, was durch die Suche mit dem Platzhalter *,?'* jedoch feststehen würde.

¹¹⁹⁵ Davies: *Corpus del Español – Género/Histórico*. Op. cit. (08.12.2011).

Ein weiteres Beispiel für die Erarbeitung von Wortfamilien anhand eines Korpus ist das Arbeitsblatt 10 – *Familias de palabras: La dictadura* (Anhang 19). Anhand dieses Arbeitsblatts soll veranschaulicht werden, dass auch im Sinne einer themenbezogenen Wortschatzerweiterung Suchen nach Wortfamilien mit dem CORPES durchgeführt werden können. Nach der Bearbeitung des Themas *La dictadura* in der Oberstufe kann somit der thematische Wortschatz zusätzlich gefestigt und erweitert werden, indem verwandte Lexeme im Korpus gesucht werden. Die Vorgehensweise ist dieselbe wie bei Arbeitsblatt 9.

Man könnte die zur selben Wortfamilie gehörigen Lexeme auch in einem Wörterbuch suchen. Der Vorteil der Korpora in Bezug auf die Erarbeitung von Wortfamilien im Vergleich zu Wörterbüchern ist jedoch eben diese Suche nach Wortstämmen mithilfe von Platzhaltern. Während man in Wörterbüchern lediglich diejenigen zu einer Wortfamilie gehörigen Lexeme findet, welche mit dem gleichen Buchstaben beginnen, kann unter Zuhilfenahme des Asterisks nach Lexemen mit dem gleichen Wortstamm, aber mit eventuellen Prä- oder Suffixen gesucht werden. So kann für das erste Beispiel auf Arbeitsblatt 9 (Anhang 18) für das Adjektiv *seguro* sowohl das Substantiv *seguridad* als auch das Verb *asegurar* gefunden werden, wohingegen man im Wörterbuch lediglich auf *seguridad* stoßen würde. Zudem stellt die Arbeit mit Korpora durch die Beschäftigung mit dem Computer einen zusätzlichen Anreiz dar. Arbeitsblätter wie dieses können zur selbstständigen Erweiterung des Wortschatzes verwendet werden, insbesondere im Hinblick auf die Textproduktion oder auf die Erweiterung von Lernstrategien. Aufgrund der relativ einfachen Eingabesprache können die Schüler schon nach wenigen gemeinsamen Suchanfragen selbstständig im Korpus nach den anderen Mitgliedern der jeweiligen Wortfamilie suchen und sie erarbeiten sich dadurch eine Möglichkeit, wie sie selbst auch zu Hause nach verwandten Lexemen suchen können.

Es sollte jedoch erneut darauf hingewiesen werden, dass manche Lexeme durch die automatische morphosyntaktische Annotation und Lemmatisierung (siehe Kapitel 1.4.5.) falsch kategorisiert werden, weshalb die Konkordanzen herangezogen werden sollten, um die Wortart im Zweifelsfall zu klären.

4.1.6. Äquivalente des deutschen „werden“

Nicht nur Kollokationen, Phraseologismen und Synonyme, sondern auch sprachliche Besonderheiten wie das deutsche ‚werden‘, welches im Spanischen auf

unterschiedliche Weise wiedergegeben werden kann, können anhand der Korpora erarbeitet werden. Als Beispiel hierfür dient das Arbeitsblatt 11 *El alemán ‚werden‘* (Anhang 20), bei dem in der ersten Aufgabe einige spanische Verben, welche dem deutschen ‚werden‘ entsprechen, in Beispielsätze eingeordnet werden sollen. Die Schüler suchen hierfür im Korpus CREA (Version 2015, siehe Kapitel 2.1.1.1.) nach einer Kombination aus dem jeweiligen Verb und dem darauffolgenden Adjektiv oder Substantiv. Wenn das passende Verb gefunden ist, kann es in der korrekten Form in den Satz eingesetzt werden. Es wurde bewusst auf die Angabe des Tempus hinter der jeweiligen Lücke verzichtet. Zum einen, da es mehrere korrekte Lösungen gibt und zum anderen, um die Aufgabe freier zu gestalten. Einige Sätze können sicherlich bereits ohne das Korpus von den Schülern mit dem korrekten Verb vervollständigt werden. Um diese Sätze zu überprüfen sowie die anderen Kombinationen herauszufinden, kann das Korpus verwendet werden. Das Arbeitsblatt kann jedoch auch dann im Unterricht genutzt werden, wenn die Schüler noch keinerlei Erfahrung mit den unterschiedlichen Übersetzungsmöglichkeiten des deutschen ‚werden‘ haben. In diesem Fall werden alle Kombinationen im Korpus nachgeschlagen. Neben der selbstständigen Zuordnung der verschiedenen Übersetzungen für das deutsche *werden*, wiederholen die Schüler mit diesem Arbeitsblatt auch die Zeitformen und die jeweiligen unregelmäßigen Verben. Für diese Teilaufgabe wird, wie bereits angegeben, das CREA (Version 2015) verwendet, weil hier ganze Phrasen eingegeben werden können. Man muss dafür allerdings einen Nachteil in Kauf nehmen: Da in diesem Korpus nicht nach Lemmata gesucht werden kann, werden nur die Treffer für die jeweils gesuchte Verbform angezeigt. Daher handelt es sich bei diesen detaillierten Suchausdrücken mit genauer Angabe eines Adjektivs oder Substantivs in einer bestimmten Form (beispielsweise *hacerse famosa*) um relativ geringe Trefferzahlen. Es kann allerdings zusätzlich nach anderen Verbformen gesucht werden, um ein besseres Bild von der Nutzung des jeweiligen Verbs in der Bedeutung ‚werden‘ zu erhalten.

In Aufgabe b) soll angegeben werden, welche Wortarten den spanischen Verben folgen, was sowohl anhand der Beispielsätze aus Übung a) als auch mit den Korpora herausgefunden werden kann. Das in dieser Aufgabe zu vervollständigende Schema

wurde in Anlehnung an Görissen (2005) und De Bruyne (2002) erstellt.¹¹⁹⁶ Die Kategorisierung befindet sich auf der zweiten Seite des Arbeitsblatts, sodass die Lehrkraft die Möglichkeit hat, zunächst nur das erste Arbeitsblatt auszuteilen, um keinerlei Informationen vorwegzunehmen und die Schüler ausschließlich mittels des Korpus zur Lösung gelangen zu lassen.

In Aufgabe c) sollen schließlich Verben für die Übersetzung deutscher Ausdrücke wie *schlechter werden*, *schlank werden* oder *Nacht werden* gesucht werden. Für diese Übung eignet sich vor allem die neue, annotierte Version des Korpus CREA, da hier auch die gesuchte Wortart (in diesem Fall: *verbo*) angegeben werden kann. Es wird somit der Wortstamm zwischen zwei Asterisken eingegeben (bei *schlechter werden* beispielsweise *peor*) und es werden die Wortart *verbo* sowie die Zeit *infinitivo* ausgewählt. Als Ergebnis werden Beispiele des Verbs *empeorar* angezeigt. Wie in Kapitel 4.1.5.1. erklärt, steht der Asterisk für ‚kein, ein oder mehrere Zeichen‘, was es dem Nutzer ermöglicht, nach optionalen Prä- oder Suffixen zu suchen. Im Fall von *empeorar* handelt es sich um ein Prä- und ein Suffix. Das Präfix vor dem Wortstamm ist jedoch nicht bei allen gesuchten Verben zu finden, beispielsweise bei *mejorar*, doch durch die Optionalität des Vorhandenseins der Prä- und Suffixe im Suchausdruck gelangt man auch hier zum korrekten Ergebnis. Auf diese Weise werden die Schüler zudem auf die Existenz von Wortfamilien aufmerksam und können bei weiterer Beschäftigung mit dem Korpus noch mehr Lexeme aus derselben Familie ermitteln. In Aufgabe d) sollen die gefundenen Verbformen aus Aufgabe c) schließlich in eigenen Sätzen verwendet werden. Um einen geeigneten Kontext für das jeweilige Verb zu finden, kann auch hier das CREA verwendet werden. Dieses Arbeitsblatt ist ab dem dritten Lernjahr geeignet.

Die Lösungen zu AB 11 *El alemán ‚werden‘* können zumindest in Teilaufgabe a) nur einen Vorschlag darstellen. Es sind auch andere Kombinationen möglich und daher im Korpus zu finden. Die Schüler sollen anhand des Korpus ermitteln, welche Verben von Muttersprachlern für gewöhnlich zum Ausdruck des deutschen ‚werden‘ im jeweiligen Kontext verwendet werden, wobei es nicht immer nur eine Lösung gibt. Diese Tatsache veranschaulicht einmal mehr, dass es in einer Sprache nicht nur eine Möglichkeit gibt, um einen Sachverhalt zu versprachlichen. Es stehen häufig parallel

¹¹⁹⁶ Vgl. Görissen, Margarita (2005): *Praxis-Grammatik Spanisch*. Stuttgart: Klett, S. 220-221. Und: De Bruyne, Jacques (2002): *Spanische Grammatik*. Tübingen: Max Niemeyer Verlag, 2. Auflage, S. 578-583.

verschiedene Formulierungsoptionen zur Verfügung, von denen manche frequenter sind als andere. Sofern das Korpus Belege für eine Formulierung bietet, in denen die Kombination auch der jeweiligen Bedeutung entspricht, können diese als Lösung akzeptiert werden.

Die Bearbeitung des Arbeitsblatts 11 zu den Entsprechungen des deutschen ‚werden‘ bereitete einigen Schülern Probleme, da sie zum einen in der Teilaufgabe a) im Korpus CREA (Version 2015) nur nach einzelnen Verbformen und nicht nach Lemmata suchen konnten, weshalb die Ergebnisse recht begrenzt waren, und es zum anderen hinsichtlich der jeweils passenden spanischen Entsprechung des deutschen ‚werden‘ teilweise verschiedene Lösungen gab. Durch die Verfügbarkeit dieser vielen Möglichkeiten waren einige Schüler überfordert. Doch nach ein paar weiteren gemeinsamen Suchanfragen und zusätzlichen Erklärungen konnten auch sie das Arbeitsblatt erfolgreich lösen. Die Teilaufgabe b) wurde aufgrund der Verunsicherung in der vorherigen Aufgabe im Klassenverband zusammen mit der Lehrkraft gelöst. Die Teilaufgaben c) und d) stellten keine großen Probleme dar.

Die Übersetzungsformen des deutschen ‚werden‘ sind ebenso gut im entsprechenden Wörterbucheintrag abzulesen, es findet sich hier jedoch in der Regel lediglich eine Liste mit den infrage kommenden Verben ohne Hinweise auf die Verwendung. Bei diesem Arbeitsblatt geht es darum, selbst eine Übersicht der verschiedenen Übersetzungsmöglichkeiten anhand der Analyse realer Sätze zu vervollständigen und sich mit der Thematik zu beschäftigen. Während in Wörterbüchern die spanischen Entsprechungen angegeben werden, können die Verben anhand der Arbeit mit Korpora in deren sprachlicher Umgebung gesehen und daraus Rückschlüsse für die Lösung der Aufgaben gezogen werden.

Um zusätzlich mehr über die Verwendungsweise spanischer Kombinationen aus Verb und Adjektiv mit der Bedeutung ‚werden‘ zu erfahren, sollten diese in einem Korpus, beispielsweise im Bereich *Colocados* des CdE:Orig (siehe Kapitel 2.1.1.3.) gesucht und analysiert werden. Zusätzlich zu AB 11 kann also mit dem CdE:Orig nach Adjektiven gesucht werden, die zusammen mit *ponerse* verwendet werden. So werden bei der Suche nach *[poner]* und *Adj.todos* im Intervall von einem Lexem rechts des Suchbegriffs die Adjektive *pálido, colorado, malo, serio, furioso, especial, junto, bueno, nervioso* und *rojo* als häufigste Treffer ermittelt. Es stellt sich somit heraus, dass *poner*

in der Regel zum Ausdruck negativer und kaum für positive Veränderungen verwendet wird, da fast ausschließlich Adjektive mit negativer Konnotation aufgelistet werden. Es finden sich zwar auch Beispiele für positive Adjektive wie *ponerse alegre / feliz / contento*, doch es überwiegen eindeutig negative Kontexte.

Diese Ergänzung der Unterrichtseinheit zu den verschiedenen Übersetzungsmöglichkeiten des deutschen ‚werden‘ kann nicht anhand eines Wörterbuchs gelöst werden, da dort neben der spanischen Entsprechung nur einzelne Beispiele für die Kombination aus Verb und Adjektiv oder Substantiv angegeben sind. Anhand der Korpora können jedoch gewöhnliche Kombinationen ermittelt werden, welche schließlich auch für die eigene Textproduktion genutzt werden können.

4.1.7. *Ser, estar und hay*

Wie Rüschhoff/Wolff bereits im Jahr 1999 bemerkten, ermöglichen Korpora, vor allem durch die Auflistung von Konkordanzen, eine schnelle Erstellung von Arbeitsblättern.¹¹⁹⁷ Ein Beispiel für ein mit Korpora erstelltes Arbeitsblatt ist AB 12 – *¿Ser, estar o hay?* (Anhang 21). Schüler haben oft Probleme damit, diese Verben zu unterscheiden und korrekt zu verwenden, weshalb sie häufig im Unterricht thematisiert werden sollten.

Für die Erstellung dieses Arbeitsblatts wurden Konkordanzen des Korpus *Sacodeyl* (siehe Kapitel 2.4.3.1.) verwendet. Aufgrund der Tatsache, dass dieses Korpus Interviews von spanischen Jugendlichen zu alltäglichen Themen wie ‚Mein Haus‘, ‚Zukunftspläne‘ oder ‚Reisen‘ beinhaltet, welche in die Niveaus A1 bis B2 des GER (siehe Kapitel 3.1.2.) eingeteilt sind, ist es thematisch sehr gut für die Erstellung von Unterrichtsmaterial geeignet und kann zudem speziell nach Texten auf einem bestimmten Niveau durchsucht werden. Im Gegensatz zu anderen Korpora können die Konkordanzen hinsichtlich der Satzlänge und Komplexität meist direkt für Übungen verwendet werden. Da es sich um ein Korpus gesprochener Sprache handelt, müssen die Sätze lediglich leicht abgeändert, teilweise Satzzeichen ergänzt und Wiederholungen gelöscht werden. Das Korpus ist sehr benutzerfreundlich, führt Suchen schnell aus und bietet mit einer Kontextlänge von maximal 15 Lexemen links und rechts des Suchausdrucks genügend Kontext, um die Beispiele für ein Arbeitsblatt zu nutzen.

¹¹⁹⁷ Vgl. Rüschhoff/ Wolff (1999): op. cit., S. 195.

Um entsprechende Beispielsätze im Korpus zu finden, wird die Funktion *Word Search* genutzt, anhand derer nach drei Lexemen gleichzeitig gesucht werden kann. Das Korpus *Sacodeyl* erlaubt leider keine Suchen nach Lemmata, sondern lediglich nach Wortformen, was bei der Erstellung dieses Arbeitsblatts jedoch kein Problem darstellt, da zum Zeitpunkt des Erlernens der Verben *ser*, *estar* und *hay* im ersten Lernjahr lediglich das Präsens von den Schülern beherrscht wird. Aufgrund dieser Tatsache muss lediglich nach 13 Formen (jeweils sechs Formen von *ser* und *estar* sowie *hay*) gesucht werden, um Sätze zu allen relevanten Wortformen zu erhalten. Sofern das Arbeitsblatt für eine Wiederholung der Grammatikeinheit zu einem Zeitpunkt genutzt wird, zu dem die Schüler bereits mehrere Zeiten gelernt haben, können auch Beispiele mit anderen Zeitformen integriert werden.

Außerdem stellt sich die Verwendung der Verben *ser* und *estar* mit einigen Adjektiven, welche je nach Kombination mit einem dieser Verben eine andere Bedeutung haben, als problematisch heraus. Eine Herangehensweise an diese Thematik stellt Arbeitsblatt 13 – *Adjetivos con ser y estar* (Anhang 22) dar. Zur Bearbeitung desselben sollen die Schüler selbstständig im CREA (siehe Kapitel 2.1.1.1.) nach den Kombinationen dieser beiden Verben mit den angegebenen Adjektiven suchen und anhand der Konkordanzen die korrekte Bedeutung ermitteln.

Um gleichzeitig nach Sätzen mit *ser* oder *estar* und dem jeweiligen Adjektiv suchen zu können, wird mittels der Plustaste eine zweite Eingabemaske geöffnet. Die beiden Eingabefelder sind als Grundeinstellung mit ‚O‘ verbunden, was eben jener Suche nach *ser* + Adjektiv ‚oder‘ *estar* + Adjektiv entspricht. Es muss also eines der beiden Kriterien erfüllt sein. Zudem wären die Optionen ‚Y‘ (beide Kriterien müssen erfüllt sein) und ‚NO‘ (das erste Kriterium muss erfüllt sein, aber nicht das zweite) möglich. In die erste Eingabemaske wird *ser* als Lemma eingegeben und mittels der Option *Proximidad* darunter das Adjektiv (beispielsweise *rico*) ebenfalls als Lemma hinzugefügt, unter der Angabe des Intervalls von einem Lexem rechts des Suchbegriffs. Ebenso wird mit *estar* in der zweiten Eingabemaske verfahren. Die Option NO kann genutzt werden, wenn beispielsweise nur Kombinationen mit *estar* unter den Konkordanzen erscheinen und man herausfinden möchte, ob es denn überhaupt solche mit *ser* im Korpus gibt. Dann kann die Option NO ausgewählt werden, um nur nach dem ersten, aber nicht nach dem

zweiten Kriterium zu suchen und somit nur Treffer mit *ser* und dem Adjektiv zu erhalten. Auf diese Weise müssen nicht alle Konkordanzen durchsucht werden, um auch Beispiele mit dem anderen Verb zu sehen.

Zusätzlich zu den obigen Einstellungen könnte die Wortart (*verbo*) angegeben werden, dies ist allerdings nicht unbedingt notwendig, da ohnehin keine anderen Wortarten angezeigt werden, mit Ausnahme von *ser* als Substantiv, welches allerdings sehr selten in Kombination mit den Adjektiven des Arbeitsblatts erscheint. Sofern die Wortart jedoch angegeben wird, werden zusätzliche Suchoptionen aufgeführt, welche die Schüler verwirren könnten. Diese zusätzlichen Angaben zu Modus, Tempus, Person und Numerus werden nicht benötigt, da ohnehin nach allen Formen der Verben *ser* und *estar* gesucht werden soll.

Sobald die Schüler die Konkordanzen zum jeweiligen Adjektiv aufgerufen haben, sollen sie die beiden angegebenen Bedeutungen der richtigen Kombination zuordnen, indem sie die Bedeutung in Kombination mit *ser* grün und diejenige mit *estar* gelb markieren. Zusätzlich sollen sie das Feld *más usado* in derjenigen Farbe markieren, deren Kombination häufiger im Korpus vorkommt. Ebenso wie bei Arbeitsblatt 7 – *Homónimos* soll auch hier die farbliche Gestaltung die Aufgabe auflockern und zur besseren Einprägung der unterschiedlichen Bedeutungen beitragen (siehe Kapitel 4.1.4.). Das Arbeitsblatt 13 ist ab Ende des zweiten Lernjahrs geeignet.

4.2. Grammatik

In Bezug auf die Nutzung von Korpora zur Erarbeitung von Grammatik im Unterricht gibt es drei verschiedene Vorgehensweisen. Zunächst können Korpora dazu dienen, um Arbeitsblätter zu erstellen, welche die Schüler (ohne Korpus) im Unterricht bearbeiten sollen. Ein Beispiel hierfür findet sich in Kapitel 4.2.6. (AB 18). Daneben können eigens für den Unterricht Arbeitsblätter mit Korpora erstellt werden, welche von den Schülern ebenso anhand der Korpora vervollständigt werden sollen, wie beispielsweise in den Kapiteln 4.2.1. (AB 14) und 4.2.2. (AB 15-17). Bei diesen beiden Vorgehensweisen werden bereits bestehende Unterrichtseinheiten um korpusbasierte Arbeitsblätter ergänzt, wobei diese im ersten Fall lediglich anhand eines Korpus erstellt wurden und im zweiten Fall zudem mittels eines Korpus bearbeitet werden sollen. Eine dritte Möglichkeit besteht darin, anhand von Konkordanzen aus Korpora Regeln abzuleiten (siehe Kapitel 4.2.5.), Grammatikeinheiten zu erarbeiten beziehungsweise zu

wiederholen (siehe Kapitel 4.2.3. und 4.2.4.) oder auch um Grammatikinhalt aus Lehrwerken zu überprüfen (siehe Kapitel 4.2.5.). In diesem Fall können sich die erarbeiteten Inhalte sogar von der bislang vermittelten Darstellung dieser Einheiten in Lehrwerken und Grammatiken unterscheiden beziehungsweise legen die entdeckten Inhalte bisweilen eine Veränderung der zu erlernenden Inhalte nahe. Bei dieser dritten Option kann die Lehrkraft den Schülern entweder zuvor aus einem Korpus ausgewählte Inhalte vorlegen oder diese selbst in einem geeigneten Korpus forschen lassen. Zu Beginn ist die vorherige Auswahl durch die Lehrkraft sicherlich zu empfehlen, doch für eine spätere Überprüfung der gelernten Inhalte anhand des Korpus finden die Schüler es unter Umständen sehr interessant, selbst im Korpus nach Belegen zu suchen.

4.2.1. Präpositionen

Schüler verwenden in ihren Textproduktionen häufig falsche Präpositionen, woraufhin die Lehrkraft diese bei der Korrektur verbessert, doch in der nächsten Textproduktion werden dieselben Fehler gemacht. Um zu erreichen, dass die Schüler die richtigen Präpositionen auswählen, kann die Lehrkraft sie diese selbst in einem Korpus ermitteln lassen. Eine Möglichkeit für dieses Vorgehen bietet Arbeitsblatt 14 – *Preposiciones* (Anhang 23), wobei die Beispiele auf diesem Arbeitsblatt beliebig ausgetauscht werden können, je nachdem, welche Präpositionen den Schülern Probleme bereiten.

In dieser Unterrichtseinheit geht es darum, zu bekannten Verben die richtige Präposition zu ermitteln. Das Arbeitsblatt umfasst hauptsächlich Verben, bei denen die Präposition nicht einfach aus dem Deutschen erschlossen werden kann oder bei denen sich die Bedeutung je nach Präposition verändert. Es empfiehlt sich, das Arbeitsblatt 14 mit dem CdE:Nuevo (siehe Kapitel 2.6.1.) zu bearbeiten, da dem Nutzer hier zunächst eine Aufzählung aller mit dem vorausgehenden Verb vorkommenden Präpositionen inklusive einer Häufigkeitsangabe angezeigt wird, woraufhin er sich die Konkordanzen für die seiner Meinung nach relevanten Präpositionen anzeigen lassen kann.

Zunächst wird also im Reiter *Colocados* das jeweilige Verb (*preocuparse*) als *Palabra/frase* eingegeben, *prep.Todos* im Bereich *Colocados* ausgewählt und das Suchintervall auf zwei Lexeme rechts des Suchbegriffs eingestellt. Nach einem Klick auf *Buscar* werden alle Präpositionen aufgelistet, die mit dem gesuchten Verb zusammen im Korpus erscheinen. Der Schüler kann sich daraufhin entweder alle Konkordanzen

anzeigen lassen oder eine Vorauswahl der Präpositionen treffen, sodass lediglich die Konkordanzen mit diesen Präpositionen aufgeführt werden. Anhand dieser Beispiele wird schließlich ermittelt, welche Präposition die richtige ist, um die deutsche Bedeutung im Spanischen zu versprachlichen. Zur Ermittlung der korrekten Präpositionen ist zu sagen, dass diese Aufgabe den Schülern äußerst leicht fällt. Das CdE:Nuevo ist sehr benutzerfreundlich gestaltet und äußerst übersichtlich, was es den Schülern erlaubt, auf einfache Art und Weise die richtigen Lösungen zu erarbeiten. Zudem hilft ihnen die Häufigkeitsangabe zu allen Präpositionen, eine Vorauswahl zu treffen.

Auf dem Arbeitsblatt ist angegeben, dass die bloße Infinitivform in die Suchmaske eingetragen werden soll. Mittels eckiger Klammern um den Suchbegriff könnte auch nach dem Lemma, also nicht nur nach dem Infinitiv, sondern auch nach konjugierten Verbformen, gesucht werden, wodurch mehr Treffer angezeigt werden würden. Bei dieser Übung ist die Suche nach der Infinitivform allerdings vorzuziehen, da die Schüler ansonsten weitere Suchoptionen beachten müssten, beispielsweise wenn nach Reflexivpronomen (*preocuparse*) oder ganzen Phrasen (*estar a favor*) gesucht werden soll. Dies birgt dennoch keinerlei Nachteil, da trotzdem genügend Beispiele im Korpus zur Verfügung stehen und die Schüler bei der Suche nach Lemmata ohnehin nicht alle Konkordanzen durchsehen könnten. Fortgeschrittene Schüler, die bereits Erfahrung mit Korpora gesammelt haben, könnten jedoch mittels korrekter Klammersetzung nach Lemmata suchen und somit mehr Beispiele zur Ermittlung der Bedeutung heranziehen.

Dieses Arbeitsblatt, das ab Ende des zweiten Lernjahrs geeignet ist, könnte auch mit anderen Korpora bearbeitet werden, die die Suche nach Wortarten zulassen, um nicht alle infrage kommenden Präpositionen zusammen mit dem Verb eingeben zu müssen. Es würden also das CREA (siehe Kapitel 2.1.1.1.) oder das CORPES (siehe Kapitel 2.1.1.2.), aber beispielsweise auch das CT-IULA (siehe Kapitel 2.5.1.1.) oder *El Grial* (siehe Kapitel 2.5.1.2.) infrage kommen. Allerdings bietet kein Korpus außer dem CdE eine Auflistung aller Präpositionen mit Frequenzangabe. Das CdE:Nuevo ist bei dieser Übung zudem besonders zu empfehlen, da es aus aktuellen Internettexten besteht und die Schüler die Bedeutung der Kombinationen aus Verb und Präposition daher aus den dort aufgelisteten Kontexten gut ableiten können. Des Weiteren ist die Eingabe in *El Grial* und im CT-IULA komplexer, da jedes Lexem des Ausdrucks einzeln eingegeben und zudem angegeben werden muss, ob es sich dabei um eine Form, ein Lemma oder eine

gesuchte Wortart handelt. Phrasen können hier nicht eingegeben werden. Bei Ausdrücken wie *estar a favor (de)* müssen somit vier *tokens* einzeln in die Suchmaske eingetragen werden, was für die Schüler relativ kompliziert ist.

Das CdE sollte allerdings nicht nur von den Schülern, sondern auch von der Lehrkraft genutzt werden, beispielsweise bei der Korrektur von Textproduktionen. Ist ihr beispielsweise nicht ganz klar, ob das Verb *tratar* nur mit der Präposition *de* oder wie in der Schülerlösung auch mit *a* verwendet werden kann, so sucht sie nach allen Formen von *[tratar]* und *prep.todos* im Umkreis von 0 Lexemen vor und 1 Lexem nach dem Verb. In diesem Fall findet das Korpus 7935 Fälle mit *de* und nur 444 Fälle mit *a*. Lässt man sich nun die Fälle mit *a* anzeigen, so ist ersichtlich, dass die Präposition hier zum Personenobjekt oder zu einer anderen Konstruktion wie *a través de* gehört. Bei der Suchmaschine *Google* hingegen werden 318,000 Treffer für *tratar a hacer* angezeigt, wodurch die Feststellung, ob es sich in diesem sprachlichen Kontext nicht doch um eine korrekt verwendete Präposition handelt, deutlich erschwert beziehungsweise der Nutzer in die Irre geführt wird. Das Korpus liefert hier eindeutige und verlässlichere Daten. Ein unklarer Fall in Bezug auf Präpositionen ist außerdem das Verb *immigrar*. Während im PONS Online-Wörterbuch *immigrar en Austria/en Suiza* als spanische Übersetzung von ‚nach Österreich/in die Schweiz immigrieren [o. einwandern]‘ angegeben wird, ist die korrekte Präposition laut der Konkordanzen des CdE:Nuevo, des CREA und des CORPES eindeutig *immigrar a*.

Korpora können im Bereich der Präpositionen also sowohl von Schülern als auch von Lehrern als Nachschlagewerk für die Korrektur von Textproduktionen, für die Suche nach Beispielen für eine bestimmte Konstruktion oder auch zur Ermittlung der Existenz von Kombinationen genutzt werden.

4.2.2. Konjugation regelmäßiger und unregelmäßiger Verben

Eine weitere Möglichkeit der Nutzung von Korpora im Spanischunterricht ist die selbstständige Erarbeitung neuer oder Wiederholung problematischer und häufig von Schülern verwechselter Verbformen unabhängig von Lehrwerkstexten. Hierfür wird ein Korpus benötigt, das neben der Wortart auch Einstellungen zu Tempus, Modus, Numerus und Person zulässt. Aus diesem Grund sind die annotierten Referenzkorpora der *Real Academia Española* (CREA und CORPES, siehe Kapitel 2.1.1.1. und 2.1.1.2.) am besten geeignet, wobei das CORPES, wie bereits in dessen Beschreibung erwähnt wurde, die

aktuelleren Texte enthält.

4.2.2.1. *Presente de indicativo*

Ein Beispiel für die Erarbeitung oder Wiederholung unregelmäßiger Verben mittels Korpora ist die Ermittlung unregelmäßiger Präsensformen anhand des CORPES. Die notwendigen Einstellungen der Suchmaske müssen lediglich einmal ausgewählt werden (*clase de palabra: verbo; modo: indicativo; tiempo: presente; persona: primera; número: singular*) und im Folgenden müssen nur noch Person und Numerus verändert werden:

The screenshot shows the CORPES search interface. At the top, there are tabs: Concordanancias, Coapariciones, Configuración, Ayuda, Modo de cita, and Sugerencias. Below the tabs, there are input fields for Lema (conocer), Forma, Clase de palabra (verbo), and Modo (indicativo). There are also checkboxes for Grafía original and a button for + Subcorpus. Below these, there are dropdown menus for Tiempo (presente), Persona (primera), Número (singular), and Tipo. At the bottom, there are buttons for Concordanancias, Estadística, and Nueva consulta. Below the search bar, it says "8.524 casos en 4.094 documentos." Below this, there is a table with columns: REF. (Clasificación, país), CONCORDANCIA, and Ordenar. The table contains 7 rows of search results.

REF. (Clasificación, país)	CONCORDANCIA	Ordenar
1 2001 Arg.	dentro de tu disco anterior (tus ídolos)? No me refiero a los músicos, sino a los otros. Conozco a algunos, pero a	
2 2001 Arg.	Charli es un genio, el único que conozco . Todo lo que hace	
3 2001 Arg.	He oído muy poco de Javier y al otro no le conozco .	
4 2001 Cuba	Perdón, no lo conozco .	
5 2001 Arg.	Si, lo conozco pero no estoy en c	
6 2001 Esp.	Conozco poco de ellos.	
7 2001 Esp.	No los conozco	

Abbildung 112: CORPES – Suche nach der 1. Person Singular von *conocer* im *presente de indicativo*.¹¹⁹⁸

Da sich das Arbeitsblatt 15 – *El presente: Formas irregulares* (Anhang 24) mit einem Grammatikkapitel beschäftigt, das bereits im ersten Lernjahr gelehrt wird, eignet es sich besser zur Wiederholung am Ende des ersten oder zu Beginn des 2. Lernjahrs als zur Erarbeitung desselben. Bei der Neudurchnahme würde die Nutzung des Korpus die Schüler verwirren und vom eigentlichen Thema ablenken. Sofern die Schüler sich zuvor mit dem CORPES beschäftigt haben, können sie es allerdings für die Wiederholung nutzen, bei der sie ihr bereits erlerntes Wissen anwenden und die restlichen Lücken mithilfe des Korpus schließen. Da den Schülern die regelmäßigen Verben am Ende des ersten Lernjahrs vertraut sein sollten, eine Wiederholung vor der Bearbeitung jedoch sinnvoll scheint, sollten die Schüler die erste Aufgabe ohne das Korpus bearbeiten – sie können es im Zweifelsfall jedoch zurate ziehen. Bei der Erstellung dieses Arbeitsblatts wurden die unregelmäßigen Verben *ser*, *estar* und *ir* absichtlich nicht mit aufgenommen, da sie den Schülern aufgrund ihres häufigen Vorkommens im täglichen Sprachgebrauch ohnehin bekannt sein dürften und eine Wiederholung daher redundant und ermüdend

¹¹⁹⁸ Real Academia Española (ed.): *CORPES. Entrada Compleja*. Op. cit. (31.08.2018).

wäre.

Da die Schüler normalerweise noch keine Erfahrung mit den Unterschieden zwischen der spanischen Sprache Lateinamerikas und der in Spanien gesprochenen haben, weil dieses Thema frühestens am Ende des 2. Lernjahrs angesprochen wird, sollte der Lehrer vor der Arbeit mit dem Korpus erwähnen, dass sich die Verbformen der 2. Person Singular im lateinamerikanischen Spanisch von denjenigen des peninsularen Spanisch unterscheiden (*pones – ponés*) und sie deshalb auf die Sprachbeispiele aus Spanien zurückgreifen sollten. Hierfür muss jedoch nicht extra in einem begrenzten Subkorpus gesucht werden, es genügt, wenn die Schüler auf die Ortsangabe auf der linken Seite der Beispiele achten.

Die angegebenen Verbformen können auch in einer Grammatik oder dem Lehrwerk nachgeschlagen werden, es soll hier aber zum einen eine Nutzung der Korpora als alternatives Nachschlagewerk vorgestellt und zum anderen eine Alternative zur lehrerzentrierten Erarbeitung im Klassenzimmer geboten werden. Es geht also darum, die gewöhnliche Grammatikeinführung im Frontalunterricht durch eine schülerzentrierte Möglichkeit des selbstständigen Forschens in realen Texten zu ersetzen, um im Unterricht zwischen verschiedenen Methoden und Sozialformen variieren zu können.

4.2.2.2. *Pretérito indefinido*

Eine weitere Möglichkeit der Nutzung von Korpora im Bereich der Flexion ist die eigenständige Erarbeitung der Formen des *pretérito indefinido*. Ein Beispiel hierfür ist das Arbeitsblatt 16 – *El pretérito indefinido* (Anhang 25). Anhand der dort vorgestellten Übungen erforschen die Schüler das CORPES und entdecken selbstständig die Formen des neu zu erlernenden Tempus. Die Schüler müssen lediglich das Verb (zum Beispiel: *continuar*) in die Suchleiste einsetzen, die Wortart (*verbo*) auswählen und *pasado* als Tempus einstellen. Danach können sie eine Form nach der anderen mittels der Angaben zu Person und Numerus suchen.

4. Nutzung von Korpora in der fremdsprachlichen Unterrichtspraxis

Concordancias Coapariciones Configuración Ayuda Modo de cita Sugerencias									
Lema	continuar	Forma		Clase de palabra	verbo	<input type="checkbox"/> Graffía original	+ Subcorpus		
Modo		Tiempo	pasado	Persona	primera	Número	singular	Tipo	
+		-		Concordancia		Estadística		Nueva consulta	
619 casos en 370 documentos.									
REF. (Clasificación, país)				CONCORDANCIA				Ordenar	
1	2001 Col.	incógnita de su adiós entre las palmeras cotidianas. Sin mover mis ojos hacia el llanto,			continué	saboreando un caj			
2	2001 Ur.	«Está con uno de sus amantes», pensé. Había perdido el control y un poco después			continué	la destrucción en e			
3	2001 Chile	inquietante, iba a abrir las puertas del pasado pero resistí, no era el momento.			Continué	por inercia, por gus			
4	2001 Chile	sentarme. La pared dejaba un hueco de medio metro, todo un mundo. Sin descansar,			continué	. La montaña segui			
5	2001 Hond.	arrastrando la pronunciación como si fuera un descubrimiento digno de pregonar, y			continué	hablándole mordaz			
6	2001 Hond.	scotch con una mujer de verdad, así que decidí relajarme y comportarme como soy, y			continué	:			
7	2001 Hond.	pensar, luego me agarre el mentón, recorrí su cuerpo de pies a ojos con la mirada y			continué	, en tono magistral:			
8	2001 Hond.	micrófono frente a mi cara, como intimidándome en respuesta a mis miradas lesivas, pero			continué	con el discurso:			
9	2001 Hond.	tan extravagantes que ningún subconsciente podría soportar tenerlos reprimidos -			continué	lentamente, levant			
10	2001 Hond.	pero cuando ella se percató de mi pausa contemplativa y se animó a decir algo, yo			continué	malcriadamente:			
11	2001 Hond.	una aparatosa victoria de los instintos sobreponiéndose a la sub cultura, así que			continué	:			
12	2001 Hond.	-Bienvenida al mundo de las cantinas bohemias -			continué	, improvisando de e			
13	2001 Méx.	aves hembras incapaces de resistir el reclamo de su macho. Con brincos en la panza			continué	con la espera que:			
14	2001 Méx.	, dejándome en plena conciencia del momento que vivía. Sigue, ordenó El Negro. Y			continué	, ya sin ti. Ahora ac			
15	2001 Méx.	indignidades de cada noche en Las Guanábanas). Esta vez tropezó con mi callada rebeldía:			continué	amamantando a m			

Abbildung 113: CORPES – Suche nach der 1. Person Singular von *continuar* im *pretérito indefinido*.¹¹⁹⁹

Eine Auswahl des Modus *indicativo* ist nicht nötig, da bei der Suche nach den Formen des *pasado*, welches ausschließlich Verben im *pretérito indefinido* herausfiltert, automatisch nur nach dem Indikativ gesucht wird (Verben im *imperfecto* oder *participio de pasado* stehen im Bereich *tiempo* ebenso als Optionen zur Auswahl und werden daher nicht als Konkordanzen bei *pasado* aufgeführt). Diese weitere Einstellung könnte die Schüler, welche sich bei der Neudurchnahme des *pretérito indefinido* erst im ersten oder zu Beginn des zweiten Lernjahrs befinden, verwirren, weshalb es ratsam ist, diese Einstellung nicht vorzunehmen. Zur Zeit der Einführung des *pretérito indefinido* sind die Spanischlerner in der Regel noch nicht mit dem *subjuntivo* in Berührung gekommen, wodurch die Unterscheidung sie nur verunsichern würde. Das Arbeitsblatt 16 eignet sich allerdings nicht nur für die Neudurchnahme, sondern auch für eine Wiederholung der Formen des *pretérito indefinido*.

Man könnte argumentieren, dass die Schüler sich die neuen Formen vor allem anhand von Ableitungsregeln erarbeiten sollten, um sie auch später neu herleiten zu können, doch gerade beim *pretérito indefinido* gibt es einige unregelmäßige Formen, die die Schüler ohnehin auswendig lernen müssen. Zudem ist es ratsam, in der Erarbeitungsweise neuer Verbformen zu variieren, um keine Langeweile bei den Schülern aufkommen zu lassen. Häufig ist die Erarbeitung neuer Zeitformen mit den dazugehörigen Unregelmäßigkeiten im Frontalunterricht eine für Schüler ermüdende

¹¹⁹⁹ Real Academia Española (ed.): *CORPES. Entrada Compleja*. Op. cit. (31.08.2018).

Angelegenheit. Die Schüler sind oft dankbar, wenn diese mühselige Phase des Lehrer-Schüler-Gesprächs durch eine aktive Erarbeitungsphase ersetzt und der Unterricht damit aufgelockert wird.

4.2.2.3. Participio de pasado

Auch die Formen des *participio de pasado* lassen sich auf einfache Weise anhand eines Korpus, beispielsweise anhand des CORPES, einführen oder wiederholen. Die Erarbeitung dieses Grammatikkapitels bietet sich bereits für Schüler an, die noch keine Erfahrung mit dem CORPES haben, da es nur weniger Einstellungen in der Suchmaske bedarf. Die Schüler müssen lediglich das jeweilige Verb (beispielsweise *hablar*) als *Lema* eingeben, die Wortart auswählen (*verbo*) und im sich daraufhin öffnenden Zusatzfeld das *participio de pasado* als *Tiempo* auswählen:

The screenshot shows the CORPES search interface. At the top, there are tabs: Concordanancias, Coapariciones, Configuración, Ayuda, Modo de cita, and Sugerencias. Below the tabs, there are input fields for Lema (hablar), Forma, Clase de palabra (verbo), Graffa original, and Subcorpus. Below these, there are dropdown menus for Modo, Tiempo (participio de pas), Persona, Número, and Tipo. At the bottom, there are buttons for Concordanancia, Estadística, and Nueva consulta. Below the search bar, it says '8.312 casos en 4.420 documentos.' Below this, there is a table with columns: REF. (Clasificación, país), CONCORDANCIA, and Ordenar.

REF. (Clasificación, país)	CONCORDANCIA	Ordenar
1 2001 Méx.	No sé si te he hablado de mi predecesor.	
2 2001 Chile	papelera. Duerme y tiene pesadillas, pero cuando despierta está seguro de que no ha hablado en sueños. Al día	
3 2001 Chile	siempre lo hubiera sido y sólo esa noche, precisamente la noche en que habíamos hablado de la muerte de la	
4 2001 Chile	cualquier cosa que me dijera iba a quedar entre nosotros dos. Dijo que siempre había hablado con franqueza cor	
5 2001 Ven.	class venezolana de aquellos años. Hoy, en pésimo estado de conservación, se ha hablado de un ambicioso p	
6 2001 Col.	provocativo y hasta mal intencionado; cabello de enredadera difícil e intrincada; hablado en un dialecto de c	
7 2001 Col.	de mi casa hasta este encierro de muerte. Tampoco sé quien me tiene ya que no he hablado con nadie, por que	
8 2001 Col.	Ha hablado de su marginalidad	

Abbildung 114: CORPES – Suche nach der Form des *participio de pasado* von *hablar*.¹²⁰⁰

Anhand des Arbeitsblatts 17 – *El participio de pasado* (Anhang 26), welches ebenso wie das Arbeitsblatt 16 ab Ende des ersten oder zu Beginn des zweiten Lernjahrs genutzt werden kann, erarbeiten die Schüler einige relevante Formen des Partizips Perfekt. Es befinden sich sowohl regelmäßig gebildete als auch unregelmäßige Partizipien darunter, wobei die Schüler in Aufgabe 2 die unregelmäßigen Formen zusätzlich farblich markieren sollen. Bei manchen Formen, wie beispielsweise dem Verb *freír* (Partizip Perfekt: *frito*), erscheint das Partizip allerdings nicht nur in der Form des Maskulinum Singular im Korpus, sondern auch als weibliche Form und im Plural. Das Auftreten dieser Formen ist der Tatsache geschuldet, dass das Partizip Perfekt *frito* als Adjektiv (beispielsweise in *patatas fritas*) verwendet wird. Die häufige Auflistung dieser Form in den Konkordanzen sollte die Schüler nicht weiter verwirren und bietet zudem

¹²⁰⁰ Real Academia Española (ed.): *CORPES. Entrada Compleja*. Op. cit. (31.08.2018).

einen guten Anknüpfungspunkt für die Lehrkraft, um den Schülern die über den bloßen Gebrauch als Teil einer Verbform hinausgehende Verwendung der Partizipien zu vermitteln.

Generell fällt die Ermittlung bestimmter Verbformen, wie sie in den Arbeitsblättern 15 bis 17 gefordert wird, den Schülern recht leicht. Aufgrund der Tatsache, dass die Einstellungen in der Suchmaske in der Regel nur einmal eingegeben und danach lediglich hinsichtlich Person und Numerus verändert werden müssen, hält sich die Zeit, welche auf komplizierte Einstellungen verwendet wird, in Grenzen und die Schüler können sich auf die Suche nach den jeweiligen Verbformen konzentrieren.

4.2.2.4. Allgemeine Nutzung von Korpora zur Ermittlung unregelmäßiger Verbformen

Wenn Schüler sich nicht sicher sind, welche die korrekte Form eines Verbs in einer bestimmten Zeit oder Person ist, können sie das CORPES heranziehen. Sie müssen lediglich das Verb und dessen Zeit sowie gegebenenfalls die Angaben zur Person in die obere Leiste eintragen. Die Schüler der 11. und 12. Klasse in den Kursen Spanisch für Spätbeginner haben häufig Probleme mit den unregelmäßigen Formen des *pretérito indefinido*. Sie müssen im 2. Lernjahr innerhalb kürzester Zeit die drei Zeiten der Vergangenheit *pretérito indefinido*, *pretérito imperfecto* und *pretérito perfecto* erlernen und haben später, wenn bereits neue Themen und Grammatikeinheiten behandelt werden, oft Schwierigkeiten, sich an die Formen zu erinnern. Ein Beispiel ist die Form der 3. Person Plural des *pretérito indefinido* von *traer*. Mit nur ein paar Klicks können sie die Form *trajeron* jedoch im CORPES finden:

The screenshot shows the CORPES search interface. At the top, there are tabs: Concordanancias, Coapariciones, Configuración, Ayuda, Modo de cita, and Sugerencias. Below the tabs, there is a search form with the following fields: Lema (traer), Forma (empty), Clase de palabra (verbo), Grafía original (checkbox), and + Subcorpus. Below this, there is a section for search criteria: Modo (indicativo), Tiempo (pasado), Persona (tercera), Número (plural), and Tipo (empty). Below the search criteria, there are buttons for Concordanancia, Estadística, and Nueva consulta. Below the search form, there is a summary: 1.869 casos en 1.374 documentos. Below this, there is a table with the following columns: REF. (Clasificación, país), CONCORDANCIA, and Ordenar por. The table contains five rows of search results, all showing the verb 'traer' in the 3rd person plural form 'trajeron'.

REF. (Clasificación, país)	CONCORDANCIA	Ordenar por:
1 2001 Esp.	mejorado desde que nombró capitán a Nasser Hussain. Hijo de hindúes, sus padres lo trajeron a este país cuando tení	
2 2001 Méx.	-Mis pies me trajeron .	
3 2001 Chile	abiertas. Celdas en donde gente a la que tú no veías espiaba todos tus movimientos. Le trajeron a un joven castrado que	
4 2001 Ven.	llevaron a sus colonias del Mediterráneo el diseño de sus ciudades, los españoles trajeron al Nuevo Mundo sus m	
5 2001 Ven.	Las lecturas ideológicas de Rousseau y de Voltaire, trajeron como consecuencia el c	

Abbildung 115: CORPES – Suche nach der 3. Person Plural des Verbs *traer* im *pretérito indefinido*.¹²⁰¹

Die Schüler werden sich nicht wegen nur einer unbekannten Verbform an den Computer setzen, doch bei der Bearbeitung der Hausaufgabe oder dem Verfassen eines

¹²⁰¹ Real Academia Española (ed.): *CORPES. Entrada Compleja*. Op. cit. (31.08.2018).

Aufsatzes ist die Nutzung des Korpus nicht abwegig. Zum einen verfassen die Oberstufenschüler ihre Aufsätze ohnehin meist am Computer, zum anderen ist die Suche im Korpus schneller als das Blättern im Lehrwerk oder der Grammatik. Zunächst müssen die Schüler im Umgang mit Korpora geschult werden, doch nach einigen Einheiten sind sie damit vertraut und die Arbeit mit Korpora stellt keine Probleme mehr dar. Sofern die Schüler einen Aufsatz verfassen, bei dem sie häufig dasselbe Tempus nur mit unterschiedlichen Verben verwenden, können die Einstellungen sogar beibehalten werden und es muss lediglich das Verb ausgetauscht werden.

4.2.3. Relativsätze

Bei der Erarbeitung oder Wiederholung bestimmter Grammatikkapitel können auch in einem Korpus ermittelte Konkordanzen gewinnbringend genutzt werden, ohne den Schülern ein Arbeitsblatt vorzulegen. Es gibt hier ebenso wie in anderen Bereichen zwei mögliche Arten, die Konkordanzen aus dem Korpus mit den Schulbuchregeln zu verbinden. Beim induktiven Lernen wird der Klasse zunächst eine Sammlung von Konkordanzen gezeigt oder die Schüler sollen selbst Beispiele für Relativsätze im Korpus suchen, sofern ihnen der Lehrer bei der Formulierung des Suchausdrucks hilft. Daraufhin sollen sie die Verwendungsweisen miteinander vergleichen, Gemeinsamkeiten und Unterschiede herausarbeiten und schließlich zu einer allgemeinen Regel gelangen. Hierbei kann die selbst erarbeitete Regel mit der Schulbuchregel verglichen und schließlich angepasst oder erweitert werden.

Bei der deduktiven Methode wird der Klasse zuerst die Schulbuchregel vorgestellt und man bespricht diese Regel sowie ihre Anwendung gemeinsam. Danach können Beispiele für Relativsätze im Korpus gesucht, miteinander verglichen und auf deren Richtigkeit hin überprüft werden. Die Klasse testet bei dieser Vorgehensweise also die zuvor eingeführte Regel anhand des Korpus, wohingegen sie bei der vorherigen Methode die Regel erst mithilfe des Korpus aufstellt. Auch Übungen, in denen die Schüler Relativsätze aus dem Korpus überprüfen sollen, welche zwar von Muttersprachlern stammen, aber nicht genau den Vorgaben der Schulbuchregel entsprechen, sind denkbar. Hier sollen die Lernenden die Abweichungen von der Regel erklären und dadurch ein besseres Gefühl für die Sprache entwickeln.

Konkordanzen können allerdings nicht nur zur Einführung der Grammatik genutzt werden, sondern auch, um zu einem späteren Zeitpunkt Grammatikeinheiten zu

wiederholen, sich die Regeln also erneut zu vergegenwärtigen. Die Arbeit mit Konkordanzen bietet somit nicht nur eine Möglichkeit der Erarbeitung bestimmter Grammatikkapitel, sondern kann ebenso der Auffrischung dienen. Es muss nicht immer ein ganzes Arbeitsblatt anhand eines Korpus bearbeitet werden, denn auch das Erschließen von Regeln anhand einiger Beispielsätze aus einem Korpus bringt Abwechslung in den Unterricht und der Lehrer stellt damit sicher, dass es sich nicht um von ihm selbst erdachte, sondern um reale Beispiele handelt.

Fortgeschrittenen Lernern ist beispielsweise manchmal nicht mehr klar, wann welche Relativpronomen verwendet werden, oder sie verwenden schlechthin keinerlei Relativsätze in ihren Textproduktionen. In diesem Fall kann die Lehrkraft den Schülern zur Wiederholung der Relativpronomen, -begleiter und -adverbien einige Konkordanzen aus dem CdE:Orig (siehe Kapitel 2.1.1.3.) vorlegen und sie die Regeln selbst ableiten lassen. Folgende Beispielsätze stammen aus dem CdE:Orig und wurden lediglich gekürzt:¹²⁰²

Ésta es la primera obra que presenta el artista en público.

El único sitio donde podía hacerlo era en Estados Unidos.

Yo soy una persona que reflexiona antes de escribir.

Hay compositores a quienes les interesa más la pintura y a otros la literatura.

Laura es una persona a la que veo a menudo.

Éramos unos ochocientos chicos de los cuales solo cien eran cristianos.

El público de España no compra un libro cuyo autor no sea conocido.

Son los músicos quienes desearían tener más tiempo para ensayar.

Anhand dieser und weiterer Konkordanzen können die Schüler die Verwendung der Relativpronomen wiederholen und die Regeln zusammen mit dem Lehrer in einem Tafelanschrieb festhalten. Die Lehrkraft kann auch weitere Konkordanzen heranziehen, das jeweilige Relativpronomen löschen und die Schüler das korrekte Pronomen ermitteln lassen. Aufgrund der Tatsache, dass beispielsweise das Relativpronomen *que* nicht nur als solches, sondern auch als Konjunktion verwendet wird, sollte mit einem morphosyntaktisch getaggten Korpus gearbeitet werden, also wie hier mit dem CdE oder mit dem CEA (siehe Kapitel 2.4.1.1.). Bei Letzterem ist das Interface zwar nicht so benutzerfreundlich wie im Fall des CdE, und die Tags der Wortarten werden nicht in der Suchmaske zur Auswahl gestellt, sondern sie müssen selbst in einem separaten *tagset*

¹²⁰² Davies: *Corpus del Español – Género/Histórico*. Op. cit. (17.02.2018).

ermittelt und eingegeben werden (Link in der Suchmaske: *Simple query language syntax*), aber es besteht zumindest die Möglichkeit, innerhalb der Option *Simple query* unter Angabe der Wortart nach einem Lexem in einer bestimmten Wortform zu suchen. In diesem Fall wäre es ‚que_PREL‘. Für andere Suchanfragen, zum Beispiel für Suchen nach Lemmata, muss allerdings die Option *CQP Syntax* ausgewählt und es muss immer eigens angegeben werden, dass nach einem Lemma gesucht wird, beispielsweise mit dem Suchausdruck [lemma=“estar“] [lemma=“verde“], um alle Formen von *estar* in Kombination mit allen Formen von *verde* als Ergebnis zu erhalten. Sollte dieses Korpus also mit Schülern genutzt werden, muss ihnen aufgrund der Komplexität der Eingabesprache genügend Zeit zur Gewöhnung an die Benutzeroberfläche gegeben werden und es müssen einige gemeinsame Suchen durchgeführt werden, damit sie mit der Benutzung des CEA nicht überfordert sind. Folgende Beispielsätze wurden dem CEA entnommen:¹²⁰³

Me referiré al discurso que el Sr Fischer pronunció la semana pasada.

Tampoco se aplica a las actividades cuyo único propósito sea la protección contra los desastres naturales.

Desde las terrazas de su segundo piso, al cual se accede por una escala de caracol, se aprecia en plenitud el morro de Arica.

En usted tenemos un experto que se ocupa con rigor de estas cuestiones.

Me sorprendió lo que escuché decir al representante de Israel, quien se refirió a las propuestas formuladas en la reunión de Ginebra.

Wie anhand dieser Konkordanzen gesehen werden kann, ist das CEA zwar aufgrund der Suchmöglichkeiten nach Lemmata und Wortarten geeignet, um nach Relativpronomen zu suchen, hinsichtlich der Texte, welche unter anderem aus dem europäischen Parlament stammen, scheint es thematisch jedoch unpassend für den Fremdsprachenunterricht.

Für Beispielsätze jeder Art ist das Korpus *Sacodeyl* (siehe Kapitel 2.4.3.1.) im Allgemeinen sehr empfehlenswert, da es sich hierbei um Interviews mit spanischen Jugendlichen handelt und die Inhalte daher sowohl thematisch als auch sprachlich für die Schüler gut zu verstehen sind. Die folgenden Relativsätze stammen aus dem *Sacodeyl*:¹²⁰⁴

No ha sido un viaje turístico como el que puedes realizar al hotel de cinco estrellas.

Mi amigo tiene una casa donde vive y otra que no usa para nada.

¹²⁰³ Subirats/ Ortega: *Corpus del Español Actual (CEA)*. Op. cit. (17.02.2018).

¹²⁰⁴ Pérez-Paredes: *Sacodeyl. European Youth Language*. Op. cit. (17.02.2018).

4. Nutzung von Korpora in der fremdsprachlichen Unterrichtspraxis

El programa cambia de nombre según la estación del año en la que estemos.

Leider kann im *Sacodeyl* lediglich nach einzelnen Wortformen und nicht nach Lemmata oder Wortarten gesucht werden. Aus diesem Grund müssen alle Formen eines Relativpronomens eingegeben werden, um Beispiele zu ermitteln (beispielsweise *el que*, *la que*, *los que* und *las que*). Zudem kann bei alleinigem Vorkommen von *que* nicht angegeben werden, dass es sich um ein Relativpronomen handeln soll, weshalb alle Konkordanzen mit *que* durchgesehen werden müssten. Die obigen Sätze sind leider einige der wenigen Relativsätze, die im Korpus zu finden sind. Aufgrund seiner überschaubaren Größe gibt es zu manchen Grammatikeinheiten nur wenige Belege, beispielsweise finden sich keinerlei Treffer zu *el/la/los/las cual/es* und *cuyo/-a/-os/-as*. Aus diesem Grund ist das Korpus *Sacodeyl*, obgleich es für viele Bereiche, beispielsweise für die Bedingungssätze im folgenden Unterkapitel, gut geeignet ist, für das Thema Relativpronomen weniger empfehlenswert.

4.2.4. Bedingungssätze

Ebenso wie das Thema der Relativpronomen können auch die Konditionalsätze anhand von Konkordanzen erarbeitet oder wiederholt werden. Folgende Konkordanzen stammen aus dem Korpus *Sacodeyl* (siehe Kapitel 2.4.3.1.):¹²⁰⁵

1. Si no tengo deberes, juego con la videoconsola.
2. ¿Qué harías si tuvieras más tiempo?
3. Por la tarde estudio si tengo exámenes o deberes.
4. Si tuvieras más tiempo libre, ¿qué actividad te gustaría hacer?
5. Si tengo un rato para descansar, descanso.
6. Si estudiara medicina, igual haría especializaciones en cirugía.
7. Si pudiera, yo estaría totalmente dispuesta a hacerlo.
8. Luego el flamenco, si me pones una sevillana, también la bailo.
9. Si en mi casa hay mucho alboroto, me voy a la biblioteca para estudiar.
10. Si pudiera o si me dieran una beca, yo lo haría.
11. Si tuvieras ocasión de irte a trabajar en el extranjero, ¿lo harías?
12. Si tuviera que elegir, me iría a Francia.

Zur Veranschaulichung der realen Bedingungssätze sowie der irrealen Bedingungssätze in der Gegenwart oder Zukunft ist das Korpus *Sacodeyl* demnach

¹²⁰⁵ Pérez-Paredes: *Sacodeyl. European Youth Language*. Op. cit. (18.02.2018).

durchaus geeignet. Die obigen Konkordanzen wurden ermittelt, indem im Bereich *Word Search* nach der Konjunktion *si* gesucht wurde.

Die Lehrkraft könnte den Schülern die obigen Konkordanzen entweder zur Erarbeitung oder zur Wiederholung vorlegen. Die beiden Typen von Bedingungssätzen werden für gewöhnlich zu unterschiedlichen Zeiten eingeführt, weshalb die obige Auflistung in der Art nicht zur Einführung der realen Bedingungssätze geeignet ist. Es könnten den Schülern jedoch zu Beginn der Unterrichtsstunde, in der dieser erste Typ zum ersten Mal thematisiert wird, die Konkordanzen 1, 3, 5, 8, 9 oder auch weitere Beispiele aus dem Korpus vorgelegt werden, mit der Frage, was ihnen bei diesen Sätzen auffällt. Auf diese Weise kann auf die Gemeinsamkeit des Vorkommens der Konjunktion *si*, deren Verwendung und Bedeutung eingegangen sowie die Zeitenfolge angesprochen werden. In der Unterrichtsstunde kann daraufhin gemeinsam ein Tafelbild erstellt sowie Übungen zum Thema bearbeitet werden. Die angegebenen Konkordanzen eignen sich außerdem zur Wiederholung der Grammatikeinheit zu Beginn der Folgestunde. In diesem Fall könnten die Verbformen ausgelassen werden, um von den Schülern vervollständigt zu werden.

Die obige Auflistung ist jedoch vielmehr für die Einführung des irrealen Bedingungssatzes in der Gegenwart gedacht. Die Lehrkraft kann anhand des Korpusmaterials auf die bereits bekannte Verwendung des ersten Typs der Bedingungssätze eingehen und die Schüler anhand der anderen Konkordanzen selbst den Gebrauch des irrealen Typs ermitteln lassen. Auch hier können die Konkordanzen als Grundlage für die Erstellung eines Tafelanschriebs dienen sowie in der Folgestunde zur Wiederholung genutzt werden. Im Falle der Wiederholung ohne Angabe der Verbformen gilt es jedoch zu beachten, dass je Satz nur eine Form gelöscht wird, da ansonsten rein aus dem sprachlichen Kontext des Satzes heraus nicht gesagt werden kann, ob es sich um den realen oder den irrealen Typ handelt.

Eine weitere Möglichkeit ist die Nutzung der Konkordanzen für die Wiederholung des Grammatikkapitels am Ende des jeweiligen Lernjahrs oder auch zu einem späteren Zeitpunkt. Die Bedingungssätze stellen in jeder Jahrgangsstufe ein heikles Thema dar, das es konstant zu wiederholen gilt. Aus diesem Grund schadet eine Auffrischung der Kenntnisse speziell vor geplanten Aufsätzen oder Schulaufgaben nicht. Die Lehrkraft kann in diesem Fall entweder der Klasse die Konkordanzen vorlegen und den Gebrauch

kurz wiederholen oder sie kann die Verbformen (lediglich eine Verbform je Satz) löschen und diese Lückenübung von den Schülern vervollständigen lassen.

Auch aus anderen Korpora können Konkordanzen zur Wiederholung der Bedingungssätze entnommen werden, doch die Themen des *Sacodeyl* sind für Schüler leicht verständlich und erleichtern dadurch die Nutzung im Unterricht. Das *Spanish in Texas Corpus* (siehe Kapitel 2.1.2.2.) und der muttersprachliche Teil des SPLLOC (siehe Kapitel 2.2.2.2.) wären von der Thematik her vergleichbar, doch in diesen beiden Korpora kann nicht nach einzelnen Lexemen gesucht, sondern es können lediglich die vollständigen Texte eingesehen werden.

4.2.5. *Subjuntivo*

Auch die Formen des *subjuntivo* könnten, ähnlich der Vorgehensweise bei den unregelmäßigen Verbformen in Kapitel 4.2.2., anhand eines Korpus erarbeitet werden, indem die Schüler das jeweilige Verb als Lemma eingeben und die gesuchte Form anhand der Optionen der Suchmaske einstellen. Mithilfe von Korpora kann allerdings auch herausgefunden werden, ob bestimmte Ausdrücke mit oder ohne *subjuntivo* verwendet werden. Dafür wird der jeweilige Ausdruck als Lemma in die Suchmaske eingegeben und anhand der Konkordanzen kann sofort erkannt werden, ob danach *subjuntivo* steht oder nicht.

Wie bereits in den Kapiteln 4.2.3. und 4.2.4. gezeigt wurde, können Konkordanzen aus einem Korpus genutzt werden, um eine Grammatikeinheit gemeinsam zu erarbeiten oder um sie zu wiederholen. Ebenso können einige Auslöser des *subjuntivo* anhand von Konkordanzen veranschaulicht werden. In diesem Fall wird das CREA (Version 2015, siehe Kapitel 2.1.1.1.) genutzt, da in der alten Version ganze Phrasen eingegeben werden können. Sollten jedoch für einen Ausdruck zu viele Konkordanzen gefunden werden (wie im Fall von *es cierto que...*), so ist es notwendig bei der Anzeige *Resultado de la consulta al banco de datos* im zweiten Auswahlbereich die Ergebnisse filtern zu lassen, um sich im nächsten Schritt die Konkordanzen anzeigen lassen zu können. Zur Veranschaulichung wurde nach den Ausdrücken *es cierto que*, *es verdad que*, *es evidente que* und *está claro que* gesucht, woraufhin folgende Konkordanzen gefunden und aus Gründen der Übersicht teilweise gekürzt wurden:¹²⁰⁶

¹²⁰⁶ Real Academia Española (ed.): *CREA (Version 2015)*. Op. cit. (19.02.2018).

Es cierto que los catalanes son muy suyos y cuando vas a Cataluña no te quieren hablar en español.

No es verdad que se hayan pagado diez millones.

Es evidente que lo estamos haciendo mal.

Es cierto que los consumidores disfrutan de alimentos importados a precios inferiores a los nacionales.

No es cierto que el cambio de ambiente haya sido beneficioso para mí.

Es evidente que esta mujer ha perdido el juicio.

Es evidente que para cada uno de los argumentos hay razones a favor y en contra.

No es mi novio y tampoco está claro que me vaya a ir a Sevilla con él.

Es verdad que cada cosa tiene su tiempo.

No es cierto que los seres humanos puedan adaptarse a cualquier medio ambiente, pero sí es verdad que el hombre ha ideado miles de formas para adaptar el medio ambiente a sus necesidades.

Está claro que los españoles (como muchos otros europeos) no piensan que la Unión Europea les pueda dar o quitar algo que les importe mucho.

Es evidente que llegamos a una situación de crisis.

Im Unterricht werden oft aus eines didaktisierten Texts, welcher die Vermittlung von *subjuntivo*-Auslösern zum Thema hat, eben solche Auslöser herausgenommen, mit den Schülern analysiert und schließlich Regeln für die Verwendung von Indikativ und *subjuntivo* nach diesen Ausdrücken aufgestellt. Anhand der vorliegenden Konkordanzen aus einem Korpus mit realen Texten ist ebenso gut zu erkennen, dass die Ausdrücke der Sicherheit *es cierto que*, *es verdad que*, *es evidente que* und *está claro que* den Indikativ nach sich ziehen. Sofern sie jedoch verneint sind, also eine Unsicherheit ausdrücken, folgt der *subjuntivo*.

Der Vorteil des CREA (Version 2015) ist, dass durch einen Klick auf den blau geschriebenen Suchbegriff ein größerer Textabschnitt um die Konkordanz herum angezeigt wird, welchen man für die Erstellung von Beispielsätzen oder auch eines Arbeitsblatts nutzen kann. Zudem ermittelt das CREA (Version 2015) die Treffer sehr schnell und ermöglicht eine repräsentative Auswahl trotz der Anwendung des Filters im Fall des Auffindens von zu vielen Konkordanzen.

Auch im opportunistischen Zeitungskorpus *Spa_news_2011_3M* (siehe Kapitel 2.1.3.1.) kann nach ganzen Phrasen gesucht werden, was ebenfalls die Suche nach Auslösern des *subjuntivo* ermöglicht. Häufig haben die Schüler Probleme mit der Verwendung des korrekten Modus nach dem Ausdruck *parece que*. Um den Gebrauch

4. Nutzung von Korpora in der fremdsprachlichen Unterrichtspraxis

von *subjuntivo* und *indicativo* nach diesem Verb zu wiederholen, können einige Konkordanzen aus diesem Korpus herangezogen werden:¹²⁰⁷

Me parece positivo que los jóvenes hagan oír su voz.

Parece que funciona bien.

La verdad es que me preocupa bastante el comportamiento del volcán, me parece que es peligroso.

No parece que usen técnicas muy diferentes a las de otros sitios.

Me parece raro que google no interceda.

Parece que cada país tiene lo que quiere por no decir lo que merece.

Me parece muy bueno que la gente progrese y que volvamos a ser exportadores de primera.

No me parece positivo que jóvenes de hoy todavía escuchen nuestra música de hace 30 años.

Todavía me parece raro que el libro exista.

No parece que ese sea un buen camino.

Parece que cada cosa que pasa nos deja en una situación peor.

No me parece que ella sea la responsable de estos resultados.

Me parece increíble que la gente diga que todos los militares violan derechos humanos.

Me parece que en el siglo XXI la armadura no va a ser en potencia militar, sino en ciencia y en tecnología.

Anhand dieser Konkordanzen kann im Unterricht erarbeitet werden, dass der Ausdruck *parece que* an sich mit dem Indikativ steht, auch in den Fällen, in denen es sich um keine objektive, sondern um eine subjektive Aussage handelt, wie bei *me parece que*. Sofern dieser Ausdruck jedoch durch ein davorstehendes *no* verneint wird, steht der *subjuntivo*. Handelt es sich um eine Bewertung der jeweiligen Aussage, welche durch ein wertendes Adjektiv zwischen *parece* und *que* ausgedrückt wird, so steht ebenfalls der *subjuntivo*.

Ein Vorteil des *Spa_news_2011_3M* ist die Tatsache, dass es sich um Zeitungstexte handelt, welche den Schülern in Bezug auf den Stil und teilweise auch hinsichtlich der Themen bekannt sind. Zudem werden die Konkordanzen zwar nicht, wie üblicherweise, so angeordnet, dass der Suchbegriff jedes Satzes zentral direkt untereinander aufgelistet ist, dafür werden allerdings vollständige Sätze angezeigt, was die Erarbeitung von Beispielsätzen erleichtert. Des Weiteren können bis zu 250 Konkordanzen auf einmal angezeigt werden.

Sowohl im CREA (Version 2015) als auch im *Spa_news_2011_3M* muss lediglich der jeweilige Ausdruck in die Benutzermaske eingetragen werden, ohne weitere

¹²⁰⁷ CLARIN (ed.): *FCS Aggregator*. Op. cit. (19.02.2018).

Einstellungen vorzunehmen. Aus diesem Grund könnte die Lehrkraft die Schüler ebenso gut selbst im Korpus suchen lassen, damit diese herausfinden, nach welchen Ausdrücken der *subjuntivo* steht. Die Lehrkraft könnte ihnen eine Reihe von Ausdrücken vorgeben, welche sie im Korpus suchen und infolgedessen angeben sollen, wann der *subjuntivo* verwendet wird.

Diese Regeln bezüglich der Verwendung von Indikativ und *subjuntivo* nach bestimmten Auslösern können auch anhand einer Grammatik erarbeitet werden, doch es soll damit eine weitere Möglichkeit gezeigt werden, wie der Unterricht mithilfe von Korpora abwechslungsreicher gestaltet werden kann und die Schüler sich die Verwendungsweise selbstständig anhand realer Sprachdaten erarbeiten können.

Korpora können außerdem, wie bereits im 3. Kapitel zu sehen war, für die Überprüfung und Verbesserung von Unterrichtsmaterialien, wie beispielsweise von Lehrwerken, dienen. Alonso Pérez-Ávila führte Untersuchungen zur tatsächlichen Verwendung von in Lehrbüchern vermittelten Ausdrücken positiver und negativer Bewertung (beispielsweise *es importante que* oder *es malo que*) durch. In diesem Kontext suchte sie die in einer Lektion des Lehrbuchs *Sueña 2* aufgeführten Ausdrücke in den zwei repräsentativen Korpora CREA (Version 2015) und CdE:Orig (Gesamt, Version 2008, siehe Kapitel 2.1.1.3. und 2.2.1.3.).¹²⁰⁸ Dabei fand sie unter anderem heraus, dass einige der im Lehrbuch angegebenen Ausdrücke nur selten oder gar nicht in diesen Korpora zu finden waren, so zum Beispiel *es útil que* und *es maravilloso que*, welche jeweils einmal im CdE:Orig (Gesamt, Version 2008) vorkamen und 14 beziehungsweise sieben Mal im CREA (Version 2015) sowie *es curativo que* und *es relajante que*, für welche es gar keine Treffer gab.¹²⁰⁹

Ähnlich der Vorgehensweise von Alonso Pérez-Ávila könnten auch Spanischlehrbücher, die an deutschen Schulen verwendet werden, im Hinblick auf den tatsächlichen Gebrauch bestimmter Ausdrücke überprüft werden. Hierbei soll ebenso wie bei Alonso Pérez-Ávila das CREA (Version 2015) verwendet werden, allerdings in Verbindung mit dem CdE:Orig, also dem 20 Mio. Lexeme umfassenden Teil des

¹²⁰⁸ Vgl. Cabrerizo Ruiz, María Aránzazu (2000): *Sueña 2*. Madrid: Anaya.

¹²⁰⁹ Vgl. Alonso Pérez-Ávila, Elena (2006): *El corpus lingüístico en la didáctica del léxico en el aula de E/LE*. Online Publikation: <https://www.mecd.gob.es/dam/jcr:c9948929-a65a-4afc-8df2-9c104e1e6175/2010-bv-11-01alonso-pdf.pdf> (06.06.2017), S. 68.

aktuellen Spanisch, da dieses ein Referenzkorpus darstellt im Gegensatz zum CdE:Orig (Gesamt), bei welchem es sich, wie in Kapitel 2.2.1.3. angegeben, um ein historisches Korpus handelt. Im Fall des CREA wird erneut die Version 2015 verwendet, da hier ganze Phrasen eingegeben werden können.

Im Lehrwerk *¡Adelante! Nivel intermedio* findet sich im Grammatikteil zu *Unidad 5* folgende Auflistung von *verbos de juicio y expresiones impersonales*, welche den *subjuntivo* nach sich ziehen:

Verb	Adjektiv	
Es Me, te... parece	bueno / malo difícil / fácil estupendo genial horrible importante interesante justo / injusto lógico necesario normal posible peligroso triste	que Alicia viaje sola.

Abbildung 116: Auflistung der *verbos de juicio y expresiones impersonales* des Lehrwerks *¡Adelante! Nivel intermedio*.¹²¹⁰

Sucht man diese Ausdrücke im CREA (Version 2015) sowie im CdE:Orig, so ergibt sich, dass besonders zwei der dort angegebenen Ausdrücke kaum oder gar nicht in den Texten der Korpora gebraucht werden, obwohl die Auflistung den Anschein erwecken könnte, dass die Ausdrücke annähernd gleichwertig in ihrer Verwendung sind. Es handelt sich dabei um die Kombinationen *es genial que*, *me/ te/ ... parece genial que*, *es estupendo que* und *me/ te/ ... parece estupendo que*. Der Ausdruck *es genial que* erzielt gar keinen Treffer in den Korpora CREA (Version 2015) und CdE:Orig, *es estupendo que* erscheint lediglich einmal im CREA (Version 2015). *Me/ te/ ... parece genial que* erscheint gar nicht im CREA (Version 2015), *me parece genial que* und *le parece genial*

¹²¹⁰ Barquero, Antonio/ Bizama, Liliana/ Corpas, Jaime/ Crovetto-Bizama, Pedro/ Díaz Gutiérrez Eva/ Jeske, Claire-Marie/ Jiménez Romera, Alicia/ Navarro, Javier/ Pardellas Velay, Rosamna (2011): *¡Adelante! Nivel intermedio. Schülerbuch*. Stuttgart: Klett, S. 161.

que erscheinen jeweils einmal im CdE und die Kombinationen mit anderen Objektpronomen kommen gar nicht vor. *Me parece estupendo* findet sich zweimal im CREA (Version 2015) und einmal im CdE, *le parece estupendo* erscheint einmal im CREA (Version 2015). Die Kombinationen mit anderen Objektpronomen erscheinen wiederum gar nicht.

Zwei weitere in dieser Grammatikeinheit angegebene Ausdrücke sind nur selten in den Korpora zu finden. *Es horrible que* erscheint lediglich acht Mal im CREA (Version 2015) und viermal im CdE:Orig. *Me parece horrible que* kommt nur einmal im CREA (Version 2015) und gar nicht im CdE:Orig vor, die anderen Formen fehlen in beiden Korpora vollständig. *Es peligroso que* erscheint zehn Mal im CREA (Version 2015) sowie dreimal im CdE:Orig und von den Kombinationen mit *parecer* erscheint lediglich einmal *me parece peligroso que* im CREA (Version 2015). Diese Werte sind vernachlässigbar gering, wenn man bedenkt, dass *es bueno que* 358 Mal im CREA (Version 2015) sowie 69 Mal im CdE:Orig erscheint und dass es für den Ausdruck *es necesario que* 1471 Treffer im CREA (Version 2015) beziehungsweise 474 Treffer im CdE:Orig gibt. Die Kombination *es posible que* kommt sogar 3122 Mal im CREA (Version 2015) und 877 Mal im CdE:Orig vor. Allerdings sind auch bei diesen drei Adjektiven die Verbindungen mit *parecer* selten. Es erscheint lediglich *me parece bueno que* fünf Mal im CREA (Version 2015) und *le parece bueno que* einmal im CdE:Orig. *Me parece necesario que* kommt dreimal im CREA (Version 2015) vor und *le parece necesario* einmal im CdE:Orig. *Me parece posible que* erscheint viermal im CREA (Version 2015), *te parece posible que* zweimal und *nos parece posible que* einmal im CdE:Orig, sonst sind diese Kombinationen nicht zu finden.

Diese Ergebnisse decken sich mit denjenigen von Alonso Pérez-Ávila, welche in ihren Untersuchungen herausfand, dass bei den Ausdrücken zur Angabe der Sicherheit ebenfalls diejenigen mit *parece* + Adjektiv + *que* (beispielsweise *parece evidente/ claro/ obvio que*) in den Korpora relativ selten vorkommen im Vergleich zu *es* + Adjektiv + *que* (beispielsweise *es cierto/ verdad/ evidente que*). Ihrer Meinung nach sollten diese Ergebnisse in irgendeiner Weise in Schulbücher oder Grammatiken aufgenommen werden, damit die Schüler die von Muttersprachlern selbst häufig gebrauchten Ausdrücke

verwenden und nicht diejenigen Kombinationen, welche selten oder kaum vorkommen.¹²¹¹

Die untersuchten Kombinationen sind in der spanischen Sprache zwar theoretisch möglich und grammatikalisch korrekt, sie kommen im tatsächlichen Sprachgebrauch jedoch nicht oder kaum vor und sind deshalb in den Korpora nicht zu finden. Folglich ist es fraglich, ob diese Ausdrücke in einem Lehrbuch für das zweite Lernjahr vermittelt oder ob sie zugunsten anderer Einheiten weggelassen werden sollten. Anstelle der Ausdrücke mit den Adjektiven *genial* und *estupendo* könnten beispielsweise *probable* und *mejor* aufgenommen werden. *Es probable que* erzielt 1040 Treffer im CREA (Version 2015) und 345 Treffer im CdE:Orig. *Es mejor que* kommt 768 Mal im CREA (Version 2015) und 169 Mal im CdE:Orig vor. Die Kombinationen mit *parecer* erscheinen in beiden Korpora jeweils gar nicht oder nur einmal. Die Ausnahme bildet *me parece mejor que*, welches acht Mal im CREA (Version 2015) erscheint.

Der hier durchgeführte Vergleich soll keine bloße Kritik an diesem speziellen Lehrwerk sein, da die Lehrbuchautoren sich aus der Fülle an Informationen auf wenige Ausdrücke beschränken müssen, um die Lerner nicht zu überfordern, und die Lektion zudem ein einheitliches Ganzes ergeben soll. Anhand dieser exemplarischen Überprüfung soll lediglich gezeigt werden, dass die Korpora der spanischen Sprache bei der Auswahl der zu vermittelnden Einheiten helfen können. Es handelt sich also vielmehr um produktive Kritik sowie um das Aufzeigen einer weiteren Möglichkeit, welche zur Verbesserung des Unterrichtsmaterials beitragen könnte.

4.2.6. Fehlerkorrektur

Auch Lernerkorpora können gewinnbringend für den Unterricht genutzt werden. Anhand von Lernerkorpora ist es unter anderem möglich, Übungen zu erstellen, in denen Fehler enthalten sind, welche die Schüler verbessern sollen. Es handelt sich hierbei um eine Überschneidung der Bereiche Lexik und Grammatik, da lexikalische und grammatikalische Fehler unterschiedlicher Art in diese Aufgaben eingehen. Derartige Übungen sind fester Bestandteil des bayerischen Jahrgangsstufentests Englisch für die 10. Klasse im Bereich *Use of English*, um das Wissen über den Gebrauch der englischen

¹²¹¹ Vgl. Alonso Pérez-Ávila (2006): op. cit. (06.06.2017), S. 66.

Sprache abzu prüfen.¹²¹² Lernerkorpora stellen solche tatsächlich von Fremdsprachenlernern gemachten Fehler bereit und ermöglichen damit die Erstellung von Übungen zur Fehlerkorrektur.

Arbeitsblatt 18 – *Corrección de errores* (Anhang 27) enthält eben solche inkorrekten Sätze aus dem Lernerkorpus CORELE (siehe Kapitel 2.2.2.1.). Das CORELE bietet die Möglichkeit der Suche nach Fehlern von Schülern mit einer bestimmten Muttersprache und ist damit besonders gut geeignet, um daraus Sätze für deutsche Spanischlerner zu entnehmen. Zudem kann unter anderem nach der Art des Fehlers und dem Sprachniveau gemäß GER (siehe Kapitel 3.1.2.) gesucht werden. Es handelt sich bei diesem Korpus zwar um ein mündliches Lernerkorpus, doch für die Fehlerkorrektur ist dies kein Problem, da sich die im Unterricht vermittelte Sprache in den ersten Jahren ohnehin mehr am mündlichen Spanisch orientiert. So heißt es beispielsweise im Jahrgangsstufenlehrplan für die 10. Klasse im Fach Spanisch als spätbeginnende Fremdsprache: „Der Schwerpunkt liegt auf Hörverstehen und Sprechfertigkeit, während die Schriftsprache zunächst eher rezeptiv einbezogen wird.“¹²¹³ Zudem wurden für das Arbeitsblatt zwar die Fehler dem Korpus entnommen, doch der sprachliche Kontext wurde vervollständigt und angepasst sowie Wiederholungen gelöscht.

Bei Arbeitsblatt 18 sollen die Schüler sich die Sätze durchlesen, Fehler markieren und diese in der Spalte auf der rechten Seite korrigieren. Ausnahmslos alle der falschen Passagen auf diesem Arbeitsblatt sind dem CORELE¹²¹⁴ entnommen und es enthält ausschließlich Fehler deutscher Spanischlerner. Bei den Sätzen des Arbeitsblatts handelt es sich um relativ einfaches Spanisch, sodass es bereits am Ende des ersten oder zu Beginn des zweiten Lernjahrs genutzt werden kann.

Ein solches Arbeitsblatt könnte auch mit anderen Lernerkorpora erstellt werden, doch das CORELE bietet den Vorteil, dass es eben nach Fehlern annotiert ist und explizit nach Fehlern von Schülern mit deutscher Muttersprache gesucht werden kann, wobei diese zudem nach verschiedenen Fehlertypen aufgelistet sind. Das SPLLOC (siehe

¹²¹² Vgl. Staatsinstitut für Schulqualität und Bildungsforschung München (ISB): *Jahrgangsstufenarbeiten (Gymnasium). Englisch*. Online Publikation: <https://www.isb.bayern.de/gymnasium/leistungserhebungen/jahrgangsstufenarbeiten-gymnasium/englisch/> (09.01.2018).

¹²¹³ Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Spanisch (spätbeginnend) – Jahrgangsstufe 10*. Op. cit. (06.09.2017).

¹²¹⁴ Universidad Autónoma de Madrid (ed.): *Corpus de Español como Lengua Extranjera (ELE)*. Op. cit. (23.02.2018).

Kapitel 2.2.2.2.), das LANGSNAP (siehe Kapitel 2.2.2.4.) und das SPT (siehe Kapitel 2.4.3.3.) verfügen lediglich über Daten englischer Muttersprachler. Zudem sind bei diesen drei Korpora zwar die Audiodateien und Transkripte verfügbar, doch es kann nicht nach Formen, Lemmata oder Wortarten gesucht werden. Das *Fono.ele* (siehe Kapitel 2.2.2.3.) würde über Daten von Spanischlernern mit deutscher Muttersprache verfügen, da es aber lediglich Audiodateien und keine Transkripte enthält, kann es nicht für die Erstellung eines solchen Arbeitsblatts genutzt werden. Das Korpus CAES (siehe Kapitel 2.2.2.5.) ist kein speziell mündliches Korpus und könnte durchaus für die Erarbeitung eines Arbeitsblatts wie AB 18 verwendet werden, da hier sogar die Suche nach Lemmata oder Wortarten zur Verfügung steht, doch es sind keine Daten zu deutschen Muttersprachlern vorhanden. Das CEDEL2 (siehe Kapitel 2.2.2.6.) ist ebenso ein Korpus geschriebener Sprache, doch hier ist weder die Suche nach Lemmata oder Wortarten möglich, noch sind Daten deutscher Muttersprachler aufgelistet. Somit ist das CORELE am besten für die Zwecke einer derartigen Übung geeignet, doch leider ist das Material dieses Korpus in den einzelnen Muttersprachen nicht besonders umfangreich, sodass es wünschenswert wäre, ein größeres Fehlerkorpus an der Hand zu haben.

4.3. Aussprache

Auch die Aussprache kann anhand von Korpora geübt und verbessert werden. Generell birgt die spanische Sprache, wie allgemein bekannt, nur wenige Schwierigkeiten in diesem Bereich. Dennoch gibt es einige Lexeme oder sprachliche Varianten, deren Aussprache mithilfe mündlicher Korpora veranschaulicht werden kann. Im CORPES (siehe Kapitel 2.1.1.2.) kann der Nutzer aus verschiedenen Subkorpora das für seine Suche passende auswählen. So ist es möglich, lediglich in mündlichen Dokumenten zu suchen. Hierfür wählt man in der Suchmaske unter *Subcorpus* die Möglichkeit *Oral* als Medium aus, wie bei der unten angegebenen Suche nach dem Ländernamen *México*, dessen Aussprache für manche Schüler eine Hürde darstellt. Durch einen Klick auf das Lautsprecher-Symbol am linken Rand der jeweiligen Konkordanz wird die Audiodatei zum ausgewählten Beispiel abgespielt:

The screenshot shows the CORPES search interface. At the top, there are tabs: ConcordanCIAS, Coapariciones, Configuración, Ayuda, Modo de cita, and Sugerencias. Below the tabs, there are search filters: Lema (México), Forma, Clase de palabra (Todos), Grafía original, Subcorpus, and Proximidad. The Subcorpus section is expanded, showing filters for Título, Autor, Fecha de clasificación, Origen (Todos, América, España, Filipinas), Medio (Todos, Escrito, Oral), Tipología (Todos, Debate, Discurso, Entrevista), Sexo (Todos, Hombre, Mujer, No indicado), and Grupo de edad (Todos, 0-14, 15-19, 20-34). There are also checkboxes for Marcas (Cursiva, Negrita, Subrayado, Versalitas). Below the filters, there are buttons for ConcordanCIA, Estadística, and Nueva consulta. The search results section shows 422 cases in 119 documents. The results are displayed in a table with columns: REF. (Clasificación, país), CONCORDANCIA, and Ordenar por: Año ascendente. The table contains five rows of search results, each showing a reference number, a year, a country, and a concordance example.

REF. (Clasificación, país)	CONCORDANCIA	Ordenar por: Año ascendente
1 2001 Perú	vez pasada le mencioné que e / i / estaba pretendiendo formar parte Érase una vez en México la película del / señor este director Robe	
2 2001 Esp.	común // es léxico común es decir utilizado / en Argentina y en y en Vitigudino // ¿eh? // en México y en Burgos / es decir que mmm la gran	
3 2001 Esp.	que tal término / es un mexicanismo / y resulta que / bueno / efectivamente se utiliza en México pero que también se está utilizando / en	
4 2001 Cuba	hispanoamericanos / que tienen ya / una formación básica en lexicografía // pienso por ejemplo en México / pienso en Colombia pienso / en Buenos	
5 2001 Esp.	entre las dos grandes lenguas de América // y // finalmente // se reintegró a su patria / México // en mil novecientos noventa y cuatro //	

Abbildung 117: CORPES – Suche nach dem Lexem *México* im mündlichen Subcorpus.¹²¹⁵

Die spanische Aussprache wird im ersten Lernjahr für gewöhnlich schrittweise durch Erklärungen der Lehrkraft eingeführt. Eine solche Einführung sollte keinesfalls durch die Arbeit mit dem Korpus ersetzt werden, allerdings haben einige Schüler auch in späteren Lernjahren trotz der überschaubaren Regeln Probleme mit der korrekten Aussprache. Teilweise vergessen sie die Regeln oder wissen bei neuem Vokabular nicht, wie dieses auszusprechen ist. Das Arbeitsblatt 19 – *La pronunciación correcta* (Anhang 28) soll daher eine abwechslungsreiche Möglichkeit der Wiederholung einiger Aussprachebesonderheiten darstellen.

Ein Thema, das häufig Probleme bereitet, ist die unterschiedliche Aussprache der Grapheme «c» und «g» vor «e/i» beziehungsweise «a/o/u» sowie die Aussprache des Graphems «u». Diese und ein paar weitere Besonderheiten sollen die Schüler anhand des CORPES erarbeiten. Die Schüler sollen hier selbst erkennen, wann ein Graphem auf welche Weise beziehungsweise ob es überhaupt ausgesprochen wird. Das Arbeitsblatt ist so konzipiert, dass die Schüler selbstständig arbeiten können, sobald sie mit dem Korpus vertraut sind und die angegebenen Symbole der Lautschrift beherrschen.

Zur phonetischen Transkription ist zu sagen, dass in der Schule eine vereinfachte Version der Lautschrift gelehrt wird, welche, in eckige Klammern gesetzt, die Aussprache neuer Lexeme im Vokabelteil des Lehrbuchs veranschaulichen soll. Um die Schüler nicht zu verwirren, bietet es sich an, auch für die Transkription der Laute auf dem Arbeitsblatt 19 diese ihnen bereits vertraute Version der Lautschrift zu übernehmen, da

¹²¹⁵ Real Academia Española (ed.): *CORPES. Entrada Compleja*. Op. cit. (31.08.2018).

derartige Übungen zur mündlichen Rezeption und Produktion weniger auf sprachwissenschaftliche Korrektheit als auf die Vermittlung muttersprachlicher Aussprache zur Nachahmung durch die Schüler abzielen.

Bei der ersten Übung auf dem Arbeitsblatt, das ab dem zweiten Lernjahr verwendet werden kann, sollen die Schüler die angegebenen Lexeme je nach der Aussprache der unterstrichenen Grapheme «c» oder «g» in eine der beiden Gruppen einteilen. Um die Lexeme richtig einordnen zu können, sollen sie im CORPES nach ihnen suchen, sich ein oder mehrere Beispiele anhören und danach ihre Entscheidung treffen. Hierbei geht es nicht nur darum, dass die Schüler die Lösung selbstständig ermitteln, sondern auch darum, dass sie eine Möglichkeit kennenlernen, wie sie zu Hause die Aussprache eines unbekannten Ausdrucks ermitteln können. In der zweiten Aufgabe geht es darum, herauszufinden, bei welchen der angegebenen Lexeme das Graphem «u» tatsächlich ausgesprochen wird und wann das Graphem «ü» stehen muss. In der letzten Übung wird schließlich auf die Aussprachebesonderheiten der Grapheme «x», «ch» und «ll» eingegangen. Am Ende jeder Aufgabe soll zudem in einer Regel die anhand der Beispiele gefundene Regelmäßigkeit festgehalten werden.

Für Fortgeschrittene bietet das Subkorpus *Oral* weitere Möglichkeiten, wie beispielsweise die Ermittlung der Ausspracheunterschiede in verschiedenen Ländern oder Regionen. Ein Beispiel ist die Ermittlung der Ausspracheunterschiede zwischen Spanien und Lateinamerika, auf die in Kapitel 4.5.3. näher eingegangen werden wird.

Die Aussprache kann prinzipiell anhand jedes Korpus des mündlichen Spanisch, das Audiodateien enthält, geübt werden. Beispiele hierfür sind das *Sacodeyl* (siehe Kapitel 2.4.3.1.), das *Backbone Spanish* (2.4.3.2.), das *Spanish in Texas Corpus* (2.1.2.2.), das *C-Or-DiAL* (2.4.2.4.), das *SPLLOC* (2.2.2.2.) oder das *LANGSNAP* (2.2.2.4.). In all diesen Korpora kann allerdings nicht speziell nach der Aussprache einzelner Lexeme gesucht werden, sondern der Nutzer kann sich lediglich die gesamte Audiodatei anhören. Die ersten drei Korpora verfügen zudem über eine schriftliche Version des jeweiligen Interviews, welche zeitgleich mitgelesen werden kann. Zu den Audiodateien des *C-Or-DiAL*, des *SPLLOC* und des *LANGSNAP* sind ebenso Transkripte verfügbar, diese sind jedoch nicht so einfach mitzulesen, da sie nicht in Sätze, sondern je nach Redepausen in einzelne Gesprächsabschnitte unterteilt sind und teilweise zusätzlich Informationen über Eigenschaften der gesprochenen Sprache enthalten. In den

Korpora *Sacodeyl* und *Backbone Spanish* kann außerdem nach einzelnen Lexemen gesucht werden, doch die anhand dieser Suche ermittelten Konkordanzen können nicht angehört werden, sondern der Nutzer müsste sich den gesamten Text anhören.

Das Korpus COLA (siehe Kapitel 2.1.2.1.) könnte allerdings für das obige Arbeitsblatt genutzt werden. Anhand der Suchmaske des COLA können Konkordanzen zu Suchbegriffen ermittelt werden, die sich der Nutzer anhören kann, ohne den gesamten Text abspielen zu müssen. Sofern es sich also auch um ein Korpus der Jugendsprache handeln darf und nicht unbedingt ein Referenzkorpus verwendet werden muss, ist das COLA durchaus für die Schulung der Aussprache zu verwenden. Es ist jedoch nicht so umfangreich wie das CORPES.

Soll also nach der Aussprache eines speziellen Lexems gesucht werden, so ist das CORPES die beste Wahl, da es sich um ein großes Referenzkorpus des Spanischen handelt, doch auch das COLA kann dafür genutzt werden. Sofern sich der jeweilige Schüler jedoch generell muttersprachliches Spanisch anhören möchte, um seine Aussprache im Allgemeinen zu verbessern, sind die Texte der anderen Korpora durchaus empfehlenswert, da es sich hier um zusammenhängende Texte unterschiedlicher Länge handelt, welche meist auch thematisch mit dem im Unterricht vermittelten Spanisch übereinstimmen.

Abgesehen vom Korpus C-Or-DiAL eignen sich die anderen Korpora der gesprochenen Sprache, welche in Kapitel 2.4.2. genauer beschrieben wurden, leider nur bedingt für die Schulung der Aussprache. Das C-ORAL-ROM könnte hierfür genutzt werden, allerdings sind die Audiodateien und Transkriptionen nur ausschnittsweise im Internet verfügbar (siehe Kapitel 2.4.2.2.). Das *Macrocorpus* ist ausschließlich auf CD-ROM verfügbar und im CORLEC sind lediglich die Transkriptionen gesprochener Texte verfügbar jedoch nicht die Audiodateien selbst.

Die Ausdrücke auf Arbeitsblatt 19 könnten sich die Schüler ebenfalls in einem Online-Wörterbuch, wie dem bereits mehrmals erwähnten Wörterbuch von PONS, vorlesen lassen.¹²¹⁶ Hier werden teilweise sogar zwei Aussprachevarianten angeboten, eine peninsulare und eine mexikanische. Allerdings findet sich die Option zum Abspielen der Audiodateien lediglich bei der Suche nach dem Pendant der deutschen Ausdrücke. Man findet die Aussprache für das Lexem *carta* demnach nur, wenn man nach dem

¹²¹⁶ Vgl. Pons (ed.): op. cit. (02.10.2018).

deutschen Lexem *Brief* sucht. Des Weiteren befindet sich nur jeweils ein Aussprachebeispiel auf der aufgerufenen Seite, sodass der Nutzer nicht mehrere zum Vergleich heranziehen kann. Dennoch kann das Arbeitsblatt 19 durchaus mit einem solchen Online-Wörterbuch bearbeitet werden, die Arbeit mit dem Korpus bietet lediglich den Vorteil, dass es die Schüler an den Umgang mit demselben gewöhnt und ihnen mehrere Beispiele beziehungsweise unterschiedliche Sprecher verschiedenen Alters und Geschlechts bietet. Das bereits erwähnte und in Kapitel 4.5.3. behandelte Arbeitsblatt zu Unterschieden im peninsularen und argentinischen Spanisch kann anhand eines solchen Online-Wörterbuchs allerdings nicht durchgeführt werden.

4.4. Textverständnis

Ein Vorteil der Existenz von Korpora für den Fremdsprachenunterricht ist die Verfügbarkeit von geschriebenen oder gesprochenen Volltexten, welche für Aufgaben zum Hör- oder Leseverstehen verwendet werden können. Leider existieren bislang nur wenige Korpora, welche Zugang zu den Volltexten gewähren. Diejenigen Korpora, welche vollständige Texte verfügbar machen, stellen jedoch eine geeignete Textquelle für Übungen zum Textverständnis dar.

4.4.1. Hörverstehen

Prinzipiell eignen sich alle Korpora des gesprochenen Spanisch, welche Audiodateien sowie im Idealfall deren Transkripte bereitstellen, für die Erstellung eines Hörverstehens. Die Texte sollten jedoch auch inhaltlich, hinsichtlich des Sprachniveaus der Schüler sowie im Hinblick auf eventuelle dialektale Färbungen passend sein, um in den Unterricht integriert werden zu können.

Das Korpus SPLLOC (siehe Kapitel 2.2.2.2.) kann beispielsweise für die Erstellung einer Hörverstehensübung verwendet werden, da es neben den Aufnahmen und Transkripten von Spanischlernern zu Vergleichszwecken auch Dateien spanischer Muttersprachler enthält. Hier kann aus verschiedenen Geschichten, Interviews und einer Paardiskussion ausgewählt werden. Auf den Aufnahmen, welche im Fall der Geschichten in der Regel zwei bis drei Minuten und im Fall der Interviews und Diskussionen ungefähr acht Minuten lang sind, finden sich allerdings einige Hintergrundgeräusche, wie das Umblättern von Seiten oder Gelächter anderer Projektteilnehmer, was die Qualität aber nicht allzu sehr beeinflusst. Das Positive an der Verwendung dieser Texte für ein

Hörverstehen im Spanischunterricht ist die Tatsache, dass es sich wirklich um gesprochene Sprache im Sinne von Koch/Österreichers konzeptioneller Mündlichkeit handelt und nicht um geschriebene Texte, die lediglich auf Band gesprochen wurden. In letzterem Fall, der im Fremdsprachenunterricht jedoch häufig anzutreffen ist, spricht man laut Koch/Österreichers lediglich von medialer Mündlichkeit.¹²¹⁷ Aus diesem Grund ist die Sprechsituation in den Audiodateien des SPLLOC durchaus als real einzustufen, da die Texte durch geringe Lexemvariation, Wiederholungen und Ausdrücke des Überlegens oder Zögerns gekennzeichnet sind. Zudem sind der Aufbau und die Abfolge der Sätze anders als in geschriebenen Texten und vermitteln den Schülern dadurch ein besseres Bild von der tatsächlichen gesprochenen Sprache.

In der für das Hörverstehen verwendeten Audiodatei mit einer Länge von etwas mehr als drei Minuten erzählt eine Muttersprachlerin eine Geschichte über zwei Schwestern, die sich bei einer Reise nach Madrid und Barcelona an ihre Kindheit erinnern.¹²¹⁸ Wie bei jedem Hörverstehen sollen sich die Schüler zunächst die Aufgaben des Arbeitsblatts 20 (Anhang 29) durchlesen und diese während des Abspielens des Hörtexts oder danach ausfüllen. Bei der ersten Aufgabe sollen die Schüler zunächst angeben, ob die Aussagen richtig oder falsch sind, und im nächsten Schritt die falschen Angaben verbessern. Danach sollen sie in einer *Multiple-Choice*-Aufgabe alle korrekten Aussagen ankreuzen, wobei es mehrere richtige Antworten geben kann. In der dritten Aufgabe soll schließlich ein Lückentext vervollständigt werden. Wie für ein Hörverstehen üblich, werden die Fragen des Arbeitsblatts der Reihe nach im Hörtext beantwortet, um die Schüler nicht zu überfordern. Das Hörverstehen kann ab dem zweiten Lernjahr verwendet werden, nachdem die Vergangenheitszeiten *pretérito indefinido* und *pretérito imperfecto* im Unterricht durchgenommen wurden, da die Geschichte in der Vergangenheit erzählt wird und aus diesem Grund die meisten Verben in diesen Zeitformen vorkommen.

In der auf der Webseite bereitgestellten Transkription des Hörtexts sind alle geäußerten Lexeme sowie unter anderem Wiederholungen und Pausen aufgeführt, doch

¹²¹⁷ Vgl. Koch/ Österreichers (2011): op. cit., S. 3.

¹²¹⁸ Vgl. Domínguez, Laura/ Mitchell, Rosamond/ Myles, Florence/ Marsden, Emma: *SPLLOC. Hermanas story. Audiofile H35NTV13N.mp3*. Online Publikation: <http://www.splloc.soton.ac.uk/data/splloc2/mp3/Hermanas/H35NTV13N.mp3> (11.01.2018).

für ein Hörverstehen im Unterricht sind diese Informationen zu detailliert.¹²¹⁹ Aus diesem Grund wurde das ursprüngliche Transkript dahingehend abgeändert, dass Wiederholungen und Reformulierungen, sofern sie nicht zum Verständnis beitragen, gelöscht und die angegebenen Pausen in Kommata und Punkte umgewandelt wurden. Das abgeänderte Transkript findet sich im Anschluss an die Lösung des Arbeitsblatts 20 in Anhang 29.

Das Korpus LANGSNAP (siehe Kapitel 2.2.2.4.), welches ebenfalls ein Lernerkorpus ist, umfasst neben den Daten der englischsprachigen Spanischlerner zu Vergleichszwecken auch Dateien von zehn Muttersprachlern. Diese Texte und Audiodateien können wie diejenigen im SPLLOC für ein Hörverstehen genutzt werden. Unter den Hörtexten befinden sich je Teilnehmer jeweils ein Interview und drei Erzählungen. In den Interviews, die jeweils ungefähr zehn Minuten lang sind, erzählen die Spanier und Mexikaner von ihrem Auslandsaufenthalt in Großbritannien, ihren Gründen dafür, den Zielen, die sie währenddessen erreichen wollen, sowie von den Erfahrungen, die sie gesammelt haben. Die Erzählungen, welche die Titel *Sisters Story*, *Cat Story* und *Brothers Story* tragen, sind zwischen zwei und drei Minuten lang. Die Hörtexte können somit ebenso gut für ein Hörverstehen genutzt werden wie die Texte des SPLLOC.

Ein Korpus, das besonders gut für die Erstellung von Aufgaben zum Hörverstehen geeignet wäre, ist das *Sacodeyl* (siehe Kapitel 2.4.3.1.). In diesem Korpus können die einzelnen Videodateien nach Themen ausgewählt werden, was die Suche nach Hörverstehenstexten, die zum aktuellen Unterrichtsthema passen, enorm erleichtert. Wie in Kapitel 3.1.3. zu sehen war, eignen sich die Interviews sehr gut für die Verwendung im Unterricht, da die Themen sich mit dem bayerischen Lehrplan überschneiden. Zudem werden die Dateien nach dem sprachlichen Niveau des GER (siehe Kapitel 3.1.2.) unterschieden, was das Verstehen seitens der Schüler gewährleistet. Die ausgewählte Datei kann entweder einem Hörverstehen dienen oder sogar einem Hör-/Schverstehen, wobei im jeweiligen Video nur die interviewte Person zu sehen ist und die Aufgaben an sich daher wie in einem gewöhnlichen Hörverstehen formuliert werden sollten.

¹²¹⁹ Vgl. Domínguez, Laura/ Mitchell, Rosamond/ Myles, Florence/ Marsden, Emma: *SPLLOC. Hermanas story. Transcript H35NTV13N.cha*.
<http://www.splloc.soton.ac.uk/data/splloc2/transcripts/Hermanas/H21NTV10N.cha> (11.01.2018).

Die Nutzung der Videodateien des Korpus *Sacodeyl* ist allerdings problematisch, da diese nur sehr abgehackt, meist aber gar nicht angesehen werden können. Die dort verfügbaren Dateien im *smil*-Format (*Synchronized Multimedia Integration Language*) können nicht mit gängigen Mediaplayern abgespielt werden. Nach Auskunft des *Sacodeyl*-Teams wird gerade an diesem Problem gearbeitet, sie geben allerdings an, dass der Fehler schwerwiegend ist und nicht so schnell behoben werden kann. Gleichzeitig verweisen sie auf das Korpus *Backbone Spanish* (siehe Kapitel 2.4.3.2.), welches ähnlich aufgebaut ist und in einer Korpusammlung der Universität Tübingen zusammen mit dem *Sacodeyl* aufgelistet wird.¹²²⁰

Anhand des Korpus *Backbone Spanish* ist es tatsächlich möglich, ein Hörverstehen zu erstellen. Es könnte auch hier ein Hör-/Sehverstehen durchgeführt werden, doch in den Videos ist lediglich der oder die Interviewte zu sehen. Die Betrachtung der sprechenden Person in ihrem Umfeld wäre für die Schüler allerdings etwas Besonderes im Gegensatz zum sonstigen Hörverstehen, welches sich lediglich auf die auditive Komponente bezieht. Zudem könnten sie auf die Artikulation des Sprechenden achten. Die Videos können im Korpus jedoch nur stockend abgespielt werden. Aus diesem Grund ist es ratsam, die Audio- oder auch Videodateien für ein Hör- oder Hör-/Sehverstehen herunterzuladen, was im Korpus *Backbone Spanish* möglich ist, um Problemen beim Abspielen vorzubeugen.

Zur Veranschaulichung wurde ein Hörverstehen zum Interview von Rocío, einer 25-jährigen Callcenter-Angestellten aus Sevilla, erstellt (siehe Arbeitsblatt 21 in Anhang 30).¹²²¹ Das Video hat insgesamt eine Länge von etwas mehr als zwölf Minuten, doch für das Hörverstehen werden lediglich die ersten vier Minuten der Aufnahme genutzt. Im ausgewählten Ausschnitt erzählt Rocío von ihrer Heimat, Sevilla, der dortigen Wirtschaft, dem Tourismus, den Sehenswürdigkeiten und traditionellen Festen, wie beispielsweise der *Semana Santa*. Dieses Interview würde sich daher für ein Hörverstehen im Rahmen einer Unterrichtseinheit zum Thema Andalusien eignen. Die einzelnen Textabschnitte sind einzeln in die Niveaus des GER eingeteilt, wobei es sich im Fall der ersten vier Minuten des Interviews hauptsächlich um Abschnitte der Niveaus A1 und A2 handelt. Somit kann das Hörverstehen bereits am Ende des ersten, vor allem aber ab dem zweiten

¹²²⁰ Vgl. E-Mail-Kontakt mit dem Ersteller-Team des Korpus *Sacodeyl* am 14.02.2018.

¹²²¹ Vgl. Kohn: *Backbone*. Op. cit. (15.02.2018).

Lernjahr genutzt werden. Die Transkripte enthalten keinerlei zusätzliche Annotationen zum gesprochenen Spanisch, weshalb sie direkt übernommen werden können und nicht mehr verändert werden müssen, im Gegensatz zu den Transkripten aus dem SPLLOC Korpus. Im Fall dieses Hörverstehens wurden lediglich die Zwischenüberschriften gelöscht. Das Transkript zum verwendeten Abschnitt der Audiodatei findet sich im Anschluss an die Lösung des Arbeitsblatts 21.

Bei der ersten Aufgabe des Hörverstehens sollen die Schüler angeben, ob die Aussagen richtig oder falsch sind, und im nächsten Schritt die falschen Angaben verbessern. Danach sollen sie drei kurze Fragen beantworten und daraufhin in einer Multiple-Choice-Aufgabe alle korrekten Aussagen ankreuzen, wobei es mehrere richtige Antworten geben kann. In der vierten Aufgabe soll schließlich ein Lückentext vervollständigt werden. Wie bereits beim Hörverstehen aus dem Korpus SPLLOC erklärt, werden die Fragen des Arbeitsblatts der Reihe nach im Hörtext beantwortet, um die Schüler nicht zu überfordern.

Das *Spanish in Texas Corpus* (siehe Kapitel 2.1.2.2.) könnte ebenfalls für Hörverstehentexte in Form von Interviews verwendet werden. Das Video oder der Hörtext kann allerdings nicht nach der Thematik der Interviews ausgewählt werden. Meist erzählen die Interviewten über ihre Vergangenheit und ihr Leben in Texas oder benachbarten Regionen. Die Aufnahmen sind mit einer Dauer von neun Minuten bis zu einer Stunde sehr lang und können daher im Unterricht lediglich in Ausschnitten für Hörverstehensaufgaben genutzt werden. Wie bereits in Kapitel 2.1.2.2. angesprochen, gibt es die Möglichkeit, sich kurze Interviewausschnitte im *SpinTX Video Archive* anzusehen, welche durchaus nach Themen ausgewählt werden können. Diese Ausschnitte sind allerdings nur von ein bis zwei Minuten Dauer und damit zu kurz für eine gewöhnliche Hörverstehensaufgabe.¹²²² Zudem gilt es zu bedenken, dass es sich bei den Interviewten um Personen handelt, die in irgendeiner Form mit dem amerikanischen Englisch in Kontakt gekommen sind, da viele von ihnen entweder in Lateinamerika geboren sind und jetzt in den USA leben, in den USA geboren sind und dann nach Lateinamerika gezogen sind oder durchgehend in den USA lebten, ihre Familie ursprünglich aber aus Lateinamerika stammt. Aus diesem Grund eignen sich die Hörtexte

¹²²² Vgl. Bullock/ Toribio (2013): *The Spanish in Texas Corpus Project*. COERLL. *SpinTX Video Archive*. Video 572. Op. cit. (06.10.2017).

beispielsweise für die Verwendung in einer Unterrichtseinheit über Mexiko, Texas oder generell über den Sprachkontakt zwischen dem Spanischen und dem Englischen.

Die Texte aus dem *Macrocorpus* (siehe Kapitel 2.4.2.1.) und dem C-ORAL-ROM (siehe Kapitel 2.4.2.2.) könnten durchaus für ein Hörverstehen genutzt werden, doch sie können aufgrund ihrer schlechten Verfügbarkeit nicht von jeder Lehrkraft ohne Weiteres verwendet werden. Im Korpus CORLEC (siehe Kapitel 2.4.2.3.) sind lediglich Transkriptionen und keine Audiodateien online verfügbar, weshalb dieses Korpus leider nicht für Hörverstehensaufgaben genutzt werden kann.

Das Korpus PRESEEA ist zwar ebenso nur auf CD-ROM verfügbar, doch wie bereits in Kapitel 2.8.2.1. erwähnt, sind einige Audiodateien und Transkripte im Internet kostenlos zugänglich. Die dort verfügbaren Dateien haben eine durchschnittliche Länge von zehn Minuten und sind durchaus für ein Hörverstehen verwendbar. Sie können nach unterschiedlichen Gesichtspunkten sortiert werden und es kann zudem nach darin enthaltenen Schlagwörtern gesucht werden. Allerdings kann nicht nach dem Thema des gesamten Interviews gesucht werden, sondern eben nur nach einzelnen Begriffen, die im jeweiligen Text enthalten sind. Die Interviews beginnen meist mit der Frage, ob die interviewte Person geduzt oder gesiezt werden möchte, und einer anschließenden Erörterung der Frage danach, wann welche Form akzeptabel ist. Für den Fall, dass die Thematik der Anrede mit ‚du‘ oder ‚Sie‘ im Unterricht besprochen wird, könnten die Texte also verwendet werden, die anderen in den Interviews enthaltenen Themen sind jedoch vorab nicht bekannt. Die Lehrkraft kann die Texte demnach nur durch das Lesen und Anhören der gesamten Interviews auf deren Nutzbarkeit im Hinblick auf den Lernstand der Schüler sowie hinsichtlich der Thematik überprüfen und auswählen. Zudem muss hinzugefügt werden, dass die Transkriptionen viele zusätzliche Angaben über den Dialog enthalten, wie Pausen, Zögern oder Unterbrechungen, welche das Lesen beträchtlich erschweren und vor einer Verwendung als Hörverstehen erst gelöscht werden müssten, um den Text als Transkription für selbiges zu verwenden.

Sofern die Lehrkraft den Schülern anhand des Hörtexts spontane gesprochene Sprache und kommunikative Funktionen vermitteln möchte, kann das Korpus C-OrDiAL (siehe Kapitel 2.4.2.4.) verwendet werden. Da bei der Aufnahme der dort verfügbaren Audiodateien viel Wert auf reale, spontane Gesprächssituationen gelegt wurde, hat der Hörer tatsächlich das Gefühl, bei einer Kommunikation im

spanischsprachigen Ausland zuzuhören.¹²²³ Anhand der Aufnahmen können daher typische Ausdrucksweisen oder kommunikative Funktionen erarbeitet werden, die einzelnen Dateien können jedoch thematisch nur begrenzt zur Erfüllung des Lehrplans dienen. Die Texte handeln zwar durchaus von Themen wie Schule, Zukunftsplänen, Freizeitgestaltung, Reisen oder Essen, doch es wird darin wenig konkreter Inhalt vermittelt – anders als in den Hörverstehenstexten, die gewöhnlich innerhalb einer Lektion verwendet werden. Vielmehr als für ein Hörverstehen nach dem obigen Beispiel (siehe Arbeitsblatt 20 und 21) würde sich eine Einheit mit diesem Korpus für die Förderung der mündlichen Ausdrucksfähigkeit und der Gesprächsführung eignen, beispielsweise vor einer mündlichen Schulaufgabe. Die Dateien in diesem Korpus können ebenso nach dem Niveau des GER sortiert werden, was die Auswahl erleichtert. Manchmal ist die Sprechgeschwindigkeit der Teilnehmer allerdings recht hoch.

Für die Nutzung der Audiodateien im Unterricht könnte die Lehrkraft den Schülern vor der Präsentation des Hörtexts zunächst allgemeine Fragen stellen, beispielsweise sollen sie das Thema herausfinden oder Schlüsselwörter heraushören. Im zweiten Durchgang könnten speziellere Aufträge gegeben werden, zum Beispiel sollen die Schüler sich typische Füllwörter zur Überbrückung von Pausen notieren. Zudem könnten sie auf Sprecherwechsel, Beginn und Ende einer Redeeinheit oder Einschübe achten.

Auf der Internetseite des Korpus stehen zu jeder Sequenz nicht nur die Audiodatei und deren reines Transkript zur Verfügung, sondern auch ein Transkript mit Angaben zu kommunikativen Funktionen.¹²²⁴ Auf diese Weise können speziell Dateien gesucht werden, die wegen der darin vorkommenden kommunikativen Funktionen für den Unterricht geeignet sind. So sollen die Schüler darauf achten, wie nach Informationen oder einer Meinung gefragt, die eigene Meinung ausgedrückt, sich nach Vorlieben und Interessen erkundigt oder etwas vorgeschlagen wird.

Die Sequenzen dieses Korpus unterscheiden sich sehr von denen anderer Korpora der gesprochenen Sprache, da sich der Nutzer aufgrund von Eigenheiten gesprochener Sprache wie Wiederholungen oder Reformulierungen in die jeweilige Situation hineinversetzt fühlt. Die Aufnahmen sind jedoch ohne die gleichzeitige Lektüre des Transkripts aufgrund des Sprechtempos, der sich überschneidenden Sprecherwechsel und

¹²²³ Vgl. Nicolás Martínez (2012): op. cit., S. 133.

¹²²⁴ Vgl. Laboratorio de Lingüística Italiana (ed.): *Acceso a las sesiones de C-Or-DiAL*. Online Publikation: <http://lablita.it/app/cordial/corpus.php> (23.02.2018).

der Kommentare der anderen Gesprächsteilnehmer teilweise schwer zu verstehen, weshalb diese Dateien, wie bereits angemerkt, nicht zur expliziten Schulung des Hörverstehens im Sinne der Ermittlung detailreicher Informationen genutzt werden können. Die Hörtexte vermitteln den Schüler allerdings ein ganz anderes, reales Bild der gesprochenen spanischen Sprache und sind daher für die Schulung der Gesprächsführung im Allgemeinen, der kommunikativen Funktionen sowie typischer Eigenschaften der gesprochenen Sprache wie Füllwörter, Überleitungen, Formulierungen zum Ausdruck der eigenen Meinung, der Zustimmung oder Ablehnung ein wertvoller Beitrag. Die Lehrkraft sollte die jeweiligen Dateien jedoch sorgfältig nach Thema, Schwierigkeitsgrad und Verständlichkeit auswählen.

Das Korpus COLA (siehe Kapitel 2.1.2.1.) könnte ebenso für die Erarbeitung von Diskursmarkern, Füllwörtern und typischen Formulierungen für Ausrufe oder Einwürfe verwendet werden. Wie das C-Or-DiAL zeichnet sich dieses Korpus durch reale gesprochene Sprache in spontanen Gesprächen aus, eignet sich daher nur bedingt zur Vermittlung relevanter Inhalte, dafür aber verstärkt für die Erarbeitung von Eigenheiten der gesprochenen Sprache. Die Dateien des Korpus COLA sind aufgrund der Tatsache, dass es sich um Audiodateien der Jugendsprache handelt, sehr interessant für ein Hörverstehen und können durchaus verwendet werden, sofern die Lehrkraft eine Hörverstehensaufgabe mit einem realen Text spanischer Jugendlicher durchführen möchte. Teilweise ist die Aufnahmequalität allerdings beeinträchtigt, da die Jugendlichen die Aufnahmegeräte bei sich tragen und diese zum einen wohl manchmal verdeckt sind, zum anderen auch viele laute Hintergrundgeräusche, wie vorbeifahrende Autos, aufzeichnen. Es sollte zudem darauf hingewiesen werden, dass diese Texte dem Hörer zwar eine gute Einsicht in die Jugendsprache und die Gesprächsführung zwischen Jugendlichen gewähren, inhaltlich jedoch nicht so viele Informationen bieten, dass ein gewöhnliches Hörverstehen erstellt werden kann, anhand dessen die Schüler spezifische Informationen herausfinden sollen. Es sollte deshalb für allgemeine Hörverstehensaufträge verwendet werden, beispielsweise könnte die Lehrkraft den Schülern ein Gespräch vorspielen und diese sollen das Thema herausfinden sowie die Meinung, welche jeder der darin vorkommenden Jugendlichen dazu hat.

4.4.2. Leseverstehen

Ebenso wie das Hörverstehen kann auch eine Übung zum Leseverstehen anhand von Texten aus Korpora durchgeführt werden. Wie bereits im 2. Kapitel angegeben, sind in vielen der spanischen Korpora leider keine vollständigen Texte verfügbar, sondern es werden lediglich Konkordanzen zum Suchbegriff angezeigt, welche teilweise um einen sprachlichen Kontext von mehreren Sätzen erweiterbar sind. Das opportunistische Zeitungskorpus *Hemero* (siehe Kapitel 2.4.1.2.) ermöglicht jedoch die Einsicht in die vollständigen Texte und kann somit als Textquelle für Aufgaben zum Leseverstehen genutzt werden.

Laut Braun/Chambers eignen sich insbesondere Zeitungskorpora für den Einsatz zu schulischen Zwecken. Zum einen ist die Textgattung für die Schüler aufgrund des hohen Bekanntheitsgrads leicht zu erschließen. Zum anderen enthalten Zeitungstexte trotz ihrer schriftlichen Form durch Zitate auch Passagen gesprochener Sprache, sodass viele Facetten der Alltagssprache repräsentiert sind.¹²²⁵

Ein Thema, welches bisweilen in der Oberstufe in Zusammenhang mit Lateinamerika behandelt wird, ist die Kinderarbeit. Wenn im Korpus *Hemero* das Schlagwort *trabajo infantil* eingegeben wird, findet das Korpus 105 passende Einträge, bei denen das Erscheinungsjahr und die Konkordanz mit dem enthaltenen Schlagwort angegeben sind sowie zwei Links, um zum vollständigen Text zu gelangen. Der erste Link zeigt den vollständigen Text innerhalb des Korpus an. Hier findet sich lediglich der Text des Zeitungsartikels unter Angabe der Internetseite, von welcher dieser stammt, es fehlen jedoch Überschriften und das Layout des ursprünglichen Zeitungsartikels. Folgt man dem zweiten Link, so wird man direkt zur entsprechenden Internetseite mit dem Originalzeitungstext geführt. Diese Aufteilung in zwei Links ist durchaus sinnvoll, denn teilweise handelt es sich um relativ kurze Zeitungstexte, welche für ein Leseverstehen auf dem Niveau der Oberstufe ungeeignet sind. Mithilfe des ersten Links kann also zunächst die Länge des Texts überprüft und der Inhalt desselben überflogen werden. Sofern der Nutzer einen Text für geeignet hält, kann er dem zweiten Link folgen und infolgedessen den Originaltext für das Leseverstehen nutzen.

Ein Beispiel für die Erstellung eines Leseverstehens ist das Arbeitsblatt 22 (Anhang 31). Der Text wurde anhand des Korpus *Hemero* gesucht und danach direkt von der

¹²²⁵ Vgl. Braun/ Chambers (2009): op. cit., S. 334.

Homepage der Zeitung *El Mundo* übernommen, wobei für die Gestaltung des Arbeitsblatts das Layout verändert wurde.¹²²⁶ Bei der ersten Aufgabe sollen die Schüler ankreuzen, ob die Aussagen korrekt oder inkorrekt sind und letztere in den Zeilen darunter verbessern. In der darauffolgenden Multiple-Choice-Übung sollen die richtigen Antworten angekreuzt werden, wobei es nicht nur eine korrekte Antwort gibt. Bei der dritten Aufgabe sollen die Schüler Fragen zum Text in vollständigen Sätzen beantworten und in der vierten Übung wird nach der Botschaft, welche die Organisation *Save The Children* innerhalb dieses Artikels vermitteln möchte, gefragt. Dieses Leseverstehen kann ab dem dritten Lernjahr durchgeführt werden.

Anders als beim Hörverstehen (siehe Kapitel 4.4.1.) sind die Lösungen der Aufgaben nicht der Reihe nach im Text zu finden. Während bei Ersterem darauf geachtet werden sollte, dass die Schüler nicht durch das Ineinandergreifen verschiedener Aufgaben unnötig verwirrt beziehungsweise überfordert werden und somit die Antworten auf die Fragen in chronologischer Reihenfolge im Hörtext vorkommen sollten, können die Aufgaben zum Leseverstehen durchaus ineinander übergehen, da der Text den Schülern vorliegt und sie die Antworten so längere Zeit vor sich haben.

Ein Vorteil der Suche von Texten zum Leseverstehen im Korpus *Hemero* im Vergleich zur bloßen Suche im Internet ist, dass es sich hierbei ausschließlich um Zeitungstexte handelt, an welche generell höhere Anforderungen bezüglich des sprachlichen und thematischen Niveaus gestellt werden können. Im Internet werden hingegen alle Texte zu einem bestimmten Thema aufgelistet und der Nutzer muss die einzelnen Seiten auf der Suche nach einem geeigneten Text selbst durchforsten. Es kann auch auf den Internetseiten der großen spanischsprachigen Zeitungen nach Texten gesucht werden, doch in diesem Fall muss bei jeder Zeitung eigens nachgeforscht werden, während das Korpus Texte aus einigen Zeitungen wie *El País*, *El Mundo* oder *El Universal* enthält und der Nutzer auf diese Weise mittels einer einzigen Suche in mehreren spanischen, argentinischen und mexikanischen Zeitungen suchen kann. Zudem ermöglicht das Korpus nicht nur die Suche nach bestimmten Lexemen oder Ausdrücken, sondern auch solche mit Platzhaltern oder nach ähnlichen Lexemen. Es können also Texte

¹²²⁶ Vgl. El mundo (2009): 'Save The Children' ve poco realista tratar de erradicar del trabajo infantil'. Online Publikation: <http://www.elmundo.es/elmundo/2009/06/10/solidaridad/1244656044.html> (17.01.2018).

4. Nutzung von Korpora in der fremdsprachlichen Unterrichtspraxis

gefunden werden, die nicht direkt das Schlagwort enthalten, sondern lediglich ähnliche Lexeme (siehe Kapitel 2.4.1.2.). Ein Nachteil ist jedoch, dass die Zeitungstexte aus den Jahren 1997 bis 2009 stammen.¹²²⁷ Diese können derzeit durchaus für ein Leseverstehen verwendet werden, doch für spätere Jahre wäre es wünschenswert, wenn neuere Texte hinzugefügt würden.

Im zweiten Korpus geschriebener Sprache, welches unter Kapitel 2.4.1. aufgelistet ist, dem CEA, sind ausschließlich Texte aus Akten, Gesetzestexten und Wikipedia-Einträgen enthalten und es sind keine vollständigen Texte verfügbar. In den beiden in Kapitel 2.1.3. vorgestellten opportunistischen Zeitungskorpora, *Spa_news_2011_3M* und *Spa_newscrawl_2011*, sind leider auch keine vollständigen Texte verfügbar. Bei Letzterem kann zudem lediglich nach einzelnen Lexemen und nicht nach mehrgliedrigen Ausdrücken wie *trabajo infantil* gesucht werden. Das CdE:Nuevo ermöglicht die Einsicht in die vollständigen Texte, allerdings nur gegen eine relativ hohe Gebühr (siehe Kapitel 2.6.1.).¹²²⁸

Das CdE:Nuevo bietet zudem ein Programm, um Texte für ein Leseverstehen vorzubereiten. Das *WordAndPhrase* aus dem Jahr 2017 ermöglicht es, ganze Texte hochzuladen und mit den Informationen aus dem Korpus abzugleichen.¹²²⁹

SAVE TEXT	FREQ RANGE	1-500	501-3000	> 3000
	411 WORDS	63 %	12 %	25 %

Inti Raymi

El Inti Raymi o Fiesta del padre Sol, **constituye** el **inicio** del año nuevo en la **cosmovisión ancestral tawantinsuyana**, **teniendo** en cuenta el Solsticio de **invierno**, del 21 de junio desde tiempos **inmemoriales**, momento en que el Sol se encuentra más **alejado** de la Alppamama o Madre tierra.

Hoy nuestro país **celebra** la " **Fiesta** del Sol " o " **Inti Raymi** ". Un **espectáculo impresionante** donde **turistas extranjeros** y **peruanos disfrutan** de la **riqueza** de nuestras **tradiciones** y **costumbres** .

¡Cusco está de **fiesta** Como cada año, la **fortaleza** de Sacsayhuamán será **escenario** de la **escenificación** del **tradicional** Inti Raymi o Fiesta del Sol, la cual se **remonta** a la época de los **incas**. Este año se espera que **lleguen** unos 3850 **espectadores**, entre **turistas extranjeros, nacionales** y **habitantes** de la zona. Lo mejor es que este año más de 380 **entradas** para ver este **espectáculo** fueron **repartidos** entre no **adultos campesinos** y aquellos que son **pacientes de hospitales**.

Pero... ¿cómo era esta **celebración** en el tiempo de los **incas**? Según el **cronista** Inca Garcilaso de la Vega, esta **fiesta** se **celebraba** siempre a **mitad** del año para **conmemorar** el origen **mítico** del Inca, según ellos, hijo del sol. A la **fiesta** **asistían** todas las personas de la ciudad de Cusco, especialmente los " **curacas** " o **jefes** de los **poblados**, quienes **traían ofrendas** para el Inca y su familia real.

WORD	PHRASE	
(CLICK TO SEE WORD SKETCH)		
LOW FREQ	MID FREQ	HIGH FREQ
2: incas, inti, ofrendas, turistas 1: abiertos, adorar, alejado, alpaca, ancestral, andas, asistían, astro, cantos, celebra, celebraba, conmemorar, consideraban, cortejo, cosechas, cosmovisión, cronista, cucillas, curacas, descaltos, dirigían, disfrutaban, disfrutaría, escenificación, espectadores, esperaban, ficticio, fortaleza, hospitales, impresionante, inmemoriales, inmenso, invocar, jefes, lleguen, mítico, ofensiva, oraciones, pagana, peruanos, poblados, rayos, realizaban, recibíéndolos, recreación, remonta, repartidos, rinde, sacrificaban, siembras, tawantinsuyana, tradiciones, traían, tributo	7: fiesta 3: llegada 2: espectáculo, extranjeros, invierno 1: adultos, brazos, campesinos, celebración, constituye, costumbres, cristiana, entradas, escenario, españoles, fueran, habitantes, histórico, inicio, mitad, nacionales, pacientes, plaza, realiza, realización, rey, riqueza,	24: de 21: la 15: del 14: y 11: el 9: en, se 8: que 7: los 6: al, año, o 5: para 4: a, esta, sol 3: con, es, este, su 2: ellos, entre, ese, espera, familia, las, luego, más, porque, según, un, una 1: ahora, animales, aquellos, buenas, cada, cerca, ciudad, como, contra, cual, cuenta, cultural, desde, después, día, donde, duda, encuentra, era, especialmente, fe, fue, fueron, hijo, hoy, junio, le, lleva, lo, mejor, mientras, momento, muy, no,

Abbildung 118: *WordAndPhrase* – Bearbeitung eines spanischen Texts zum Fest *Inti Raymi*.¹²³⁰

¹²²⁷ Vgl. Calvo, José (2016): *Textdaten Romanistik. Corpus periodístico Hemero*. Op. cit. (14.03.2017).

¹²²⁸ Vgl. Davies: *Full-text corpus data from corpus.byu.edu – Purchase data*. Op. cit. (13.12.2016).

¹²²⁹ Vgl. Davies: *WordAndPhrase*. Op. cit. (20.01.2018).

¹²³⁰ Davies, Mark: *WordAndPhrase. Analyze text*. Online Publikation: <https://www.wordandphrase.info/span/x.asp> (20.01.2018).

In dieses Programm kann jeder beliebige Text hochgeladen und analysiert werden. Im oben angegebenen Beispiel handelt es sich um einen Text zum Fest *Inti Raymi*, welcher beispielsweise innerhalb der Unterrichtseinheit *Perú* zu Beginn des zweiten Lernjahrs Spanisch als spätbeginnende Fremdsprache im Unterricht besprochen werden kann.¹²³¹ Auf der linken Seite werden die einzelnen Lexeme je nach deren Frequenz im Korpus in unterschiedlichen Farben angezeigt und auf der rechten Seite findet sich eine Auflistung aller Lexeme unter Angabe von deren Häufigkeit im Text. Mittels eines Klicks auf eines der Lexeme springt das Programm zum entsprechenden Eintrag, in welchem Definitionen, Synonyme, Kollokationen und relevante Themen angegeben sind:

REPARTIR v (RANK 2529, FREQ 58364)

☆ 🔄 🗑️ **repartir**, poblado, dios, altar, ofrenda, ofrenda, invocar, poblado, repartir, repartir

DEFINITIONS (click for more)
give out deal deliver share out.
distribute hand out divide up (issue)

SYNONYMS (click to see word)
disregar v (scatter) dispersar, esparcir, desparramar, propalar, sembrar, irradiar, diseminar,
distribuir v (allocate) prorratar, ratear,
dividir v (split) distribuir, racionar, asignar, donar, adjudicar, dosificar,
graduar v (graduate) medir, determinar, dividir, administrar, tomar, dosificar,
medir v (survey) tasar, limitar, restringir, racionar,
mezclar v (blend) revolver, cortar, barajar,
nivelar v (cover) dividir, igualar, equivaler, distribuir, dosificar, promediar,
seccionar v (dissect) dividir, separar, cortar, desmenuzar, fragmentar, descomponer, fraccionar,

Corrections / problems ?

COLLOCATES ○ new word ● with REPARTIR
(NOUN) dinero, riqueza, volante, comida, beneficio, partes, botín, premio, dividendo, pastel, bienes, folleto, pan, condón, regalo, pobre, herencia, utilidad, torta, poesía
(MISC) equitativamente, repartir, equitativo, restante, gratis, informativo, gratuitamente, inscrito, vano, proporcionalmente, uniformemente, desigualmente, siniestro, doquier, uniforme, generosamente, desigual, gratis, proporcional, fraccionar

TOPICS (click to see)
reparto n euro n castro n castrista n dividendo n fidel n renta n castrismo n n, indio n revolucion n castrista n accionista n discípulo n testamento n capitalismo n apóstol n pan n siervo n esbirro n bienes n jugador n salario n carta n herencia n oligarquía n tiranía n pai n riqueza n chivato n

Abbildung 119: *WordAndPhrase* – Eintrag des Verbs *repartir*.¹²³²

Sollte die Lehrkraft einige Lexeme des Texts ersetzen wollen, beispielsweise weil sie sich wiederholen oder weil sie den Schülern bislang unbekannt sind und auch nicht eingeführt werden sollen, so kann sie anhand dieses Programms nach Synonymen suchen. Sollte unbekanntes Vokabular im Originaltext erhalten bleiben, so können die Synonyme auch für Vokabelangaben verwendet werden. Im Text über *Inti Raymi* sind einige Lexeme enthalten, welche den Schülern im zweiten Lernjahr noch nicht bekannt sind. Im

¹²³¹ Der im Unterricht verwendete Text wurde auf der Grundlage der folgenden Internetquellen erstellt: Diario Correo (ed.): *Cusco celebra hoy escenificación de tradicional Inti Raymi*. Online Publikation: <https://diariocorreo.pe/ciudad/cusco-celebra-hoy-escenificacion-de-tradicional-inti-raymi-527502/> (31.10.2016).

El montonero (ed.): *El Inti Raymi y el Inca Garcilaso de la Vega*. Online Publikation: <http://elmontonero.pe/cultura/el-inti-raymi-y-el-inca-garcilaso-de-la-vega> (31.10.2016).

Vega, Juan José/ Guzmán Palomino, Luis (2005): *El Inti Raymi Inkaico. La verdadera historia de la gran fiesta del sol*. Online Publikation:

http://sisbib.unmsm.edu.pe/Bibvirtual/Publicaciones/antropologia/2003_N01/a05.htm#* (31.10.2016).

¹²³² Davies, Mark: *WordAndPhrase. Repartir*. Online Publikation:

<https://www.wordandphrase.info/span/x3.asp?d=y&w1=repartir&c1=v> (20.01.2018).

4. Nutzung von Korpora in der fremdsprachlichen Unterrichtspraxis

Folgenden sind vier dieser Lexeme zusammen mit Synonymen aufgelistet, welche dem jeweiligen Eintrag im Programm *WordAndPhrase* entnommen wurden und für Vokabelerklärungen genutzt werden könnten:

repartir algo = distribuir
un poblado = un pueblo
una ofrenda = un regalo, un sacrificio
invocar a alguien = rogar, orar

Abbildung 120: Vokabelerklärung für vier unbekannte Lexeme aus dem Text *Inti Raymi*.

Im Anschluss an Aufgaben zum Leseverstehen werden häufig Lückenübungen durchgeführt, in denen die Schüler das zu erlernende Vokabular in einen neuen sprachlichen Kontext einbauen sollen. Die Lehrkraft kann auf der Suche nach Beispielsätzen entweder die Konkordanzen im jeweiligen Eintrag verwenden oder im Bereich *Collocates* die Option *with XY* (in Abbildung 119 *with REPARTIR*) auswählen, um mittels eines Klicks auf das entsprechende Lexem (zum Beispiel *dinero*) zu den Konkordanzen dieser Kollokation zu gelangen. Für die angegebenen vier neuen Vokabeln könnten folgende Sätze aus den Konkordanzen in eine Lückenübung eingebaut werden:

1. Anabella _____ globos a sus amiguitos en una fiesta. (→ reparte)
2. Las ciudades pequeñas y los _____ a lo largo de todo el Ecuador son muy tranquilos. → (poblados)
3. La estatua de Neb Senu, que tiene una altura de aproximadamente 25 centímetros, es una _____ al dios egipcio Osiris. (→ ofrenda)
4. Oxfam _____ al Gobierno peruano a buscar una solución. (→ invoca)

Abbildung 121: Lückenübung für vier unbekannte Lexeme aus dem Text *Inti Raymi*.

Im Programm *WordAndPhrase* kann sich der Nutzer allerdings nicht nur die Einträge einzelner Lexeme anzeigen lassen, sondern es ist zudem möglich, nach ganzen Phrasen im Korpus zu suchen. Hierfür wird im Reiter *Analyze Text* anstatt der Grundeinstellung *Word* im rechten oberen Eck die Einstellung *Phrase* ausgewählt.

Daraufhin können bis zu fünf Lexeme des hochgeladenen Texts ausgewählt und im CdE:Nuevo ermittelt werden. Es kann wortwörtlich nach Beispielen für die eingegebene Phrase gesucht werden oder nach Synonymen für bestimmte Lexeme. Zudem ist die Suche mit Platzhaltern, Lemmata oder Wortarten möglich. Beispielsweise kann eine Kombination aus zwei Lexemen aus dem hochgeladenen Text ausgewählt und Synonyme für diese Konstellation ermittelt werden. Diese Funktion ist auch für die Textproduktion nicht zu verachten, da nicht nur nach Synonymen für einzelne Lexeme, sondern auch solche für ganze Gruppen von Lexemen gesucht werden können.

Aufgrund der soeben besprochenen Möglichkeiten bietet es sich an, dieses Programm vor allem Schülern in fortgeschrittenen Lernjahren vorzustellen. Es kann von ihnen zur Verbesserung ihrer Textproduktionen verwendet werden, indem sich wiederholende Lexeme erkannt werden und anhand des Korpus nach Synonymen gesucht werden kann. Zudem geben die Einträge zu jedem einzelnen Lexem einen guten Überblick über dessen Bedeutung, Kollokationen und typischen sprachlichen Kontext.

Ein weiteres Programm, welches für die Vorbereitung eines Leseverstehens verwendet werden kann, ist das *Wordlist Tool* des in Kapitel 2.9. vorgestellten *WebCorp Live*.¹²³³ In dieses Programm kann eine beliebige Internetadresse eingegeben und eine *Wordlist* des Inhalts der Seite erstellt werden. Vor der Bearbeitung des Themas *Alhambra* kann beispielsweise zum Inhalt einer Internetseite zum *Generalife* eine *Wordlist* erstellt werden, anhand derer relevantes, vorab zu semantisierendes Vokabular ermittelt werden kann.¹²³⁴ Für derartige Suchen ist es sinnvoll, nicht zwischen Groß- und Kleinschreibung zu unterscheiden sowie den Filter für *Stopwords* zu verwenden, um frequente Funktionswörter wie Artikel von vornherein auszuschließen. Auf der verwendeten Webseite zum Thema *Generalife* ist das häufigste Lexem *alhambra* (50 Treffer), gefolgt von *generalife* (24), *patio* (12), *jardines* (11), *granada* (10), *sala* (10), *paseo* (10) und *museo* (10). Außerdem erscheinen Lexeme wie *visitantes* (9), *palacio* (8), *reyes* (7), *cipreses* (7) oder *restauración* (6) weit oben in der Liste.¹²³⁵ Zudem kann die N-Gramm-

¹²³³ Vgl. Research and Development Unit for English Studies, Birmingham City University (ed.): *WebCorp Live. Concordance the web in real-time. Wordlist Tool*. Online Publikation: <http://www.webcorp.org.uk/live/wdlist.jsp> (08.12.2016).

¹²³⁴ Vgl. Junta de Andalucía (ed.): *Conocer la Alhambra. Generalife*. Online Publikation: <http://www.alhambra-patronato.es/index.php/Generalife/31/0/> (08.12.2016).

¹²³⁵ Vgl. Research and Development Unit for English Studies, Birmingham City University (ed.): *WebCorp Live. Concordance the web in real-time. Wordlist einer Internetseite zum Generalife mit N-*

Größe zwischen 1 und 5 variiert werden (siehe auch Kapitel 1.3.2.1.). Obige *Wordlist* erscheint bei der Größe 1. Sofern man die Größe 5 auswählt, finden sich die häufigsten Kombinationen mit fünf Lexemen in der Liste, wie beispielsweise *la alhambra y el generalife* (6), *la sala de los reyes* (3) oder *el museo de la alhambra* (2).¹²³⁶ Die auf diese Weise ermittelten, für das Verständnis des jeweiligen Texts ausschlaggebenden Lexeme können auf ihre Bekanntheit hin überprüft werden und unbekannte Ausdrücke entweder vorab eingeführt oder als Fußnote angegeben werden.

4.5. Landeskunde

Landeskundliche Themen des Spanischunterrichts können nur bedingt mit Korpora bearbeitet werden, doch es ist möglich, sie mit deren Hilfe zu ergänzen. So können Texte aus Korpora, wie beispielsweise aus dem Zeitungskorpus *Hemero* (siehe Kapitel 2.4.1.2. sowie 4.4.2.), dem Korpus *Sacodeyl* (2.4.3.1.), dem *Spanish in Texas Corpus* (2.1.2.2.) oder auch dem CEDEL2 (2.2.2.6.), im Unterricht verwendet werden, um landeskundliche Informationen zu vermitteln. Während das *Hemero* Zeitungstexte zu verschiedenen Themen bietet, können anhand des *Sacodeyl* und des *Spanish in Texas Corpus* Informationen zum Leben spanischer Jugendlicher beziehungsweise spanischer und bilingual spanisch-englischer Texaner gesammelt werden, wie es in Kapitel 4.5.1. der Fall ist. Auch das Lernerkorpus CEDEL2 bietet muttersprachliche Texte, welche für die Vermittlung landeskundlicher Inhalte genutzt werden können. Des Weiteren ist es möglich, anhand von Korpora sprachliche Aspekte, welche in Zusammenhang mit einem bestimmten Thema stehen, zu veranschaulichen, wie in den Kapiteln 4.5.2., 4.5.3. und 4.5.4. zu sehen sein wird.

4.5.1. Alltag und Hobbies spanischer Jugendlicher

Ein Themengebiet, welches durchaus anhand eines Korpus erarbeitet oder vertieft werden kann, ist das Leben spanischer Jugendlicher. Wie bereits in den Kapiteln 2.4.3.1. und 3.1. erwähnt, enthält das Korpus *Sacodeyl* Interviews spanischer Jugendlicher zu

Gramm-Größe 1. Online Publikation:

<http://www.webcorp.org.uk/live/wdlist.jsp?url=http%3A%2F%2Fwww.alhambra-patronato.es%2Findex.php%2FGeneralife%2F31%2F0%2F&sw=on&i=on&n=1> (08.12.2016).

¹²³⁶ Vgl. Research and Development Unit for English Studies, Birmingham City University (ed.): *WebCorp Live. Concordance the web in real-time. Wordlist einer Internetseite zum Generalife mit N-Gramm-Größe 5*. Online Publikation:

<http://www.webcorp.org.uk/live/wdlist.jsp?url=http%3A%2F%2Fwww.alhambra-patronato.es%2Findex.php%2FGeneralife%2F31%2F0%2F&sw=on&i=on&n=1> (08.12.2016).

ihrem Alltag sowie zu Hobbies, Ferien, Schule und ihren Zukunftsplänen. Meist werden derartige Themen im Unterricht in Form von Lehrbuchtexten als Hör- oder Leseverstehen behandelt. Die Schüler erarbeiten sich somit anhand eines Texts aus dem Lehrbuch die Lebensweise spanischer Jugendlicher, sprechen infolgedessen darüber und vergleichen deren Alltag mit dem eigenen. Anhand des Korpus *Sacodeyl* stehen der Lehrkraft und den Schülern zusätzliche Texte spanischer Jugendlicher zur Verfügung, die sie in den Unterricht einfließen lassen können. Derartige Zusatztexte können aber auch als Grundlage für Schulaufgaben dienen.

Wie in Kapitel 3.1.3.3. zu sehen war, sind der Alltag und die Freizeitgestaltung der Jugendlichen Thema im zweiten Lernjahr Spanisch als dritte Fremdsprache. Da die Schüler für gewöhnlich großes Interesse am Leben Gleichaltriger in anderen Ländern haben, ist es denkbar, in der neunten Klasse eine Unterrichtseinheit einzubauen, in welcher die Schüler ihren Klassenkameraden den Alltag eines spanischen Jugendlichen präsentieren sollen. Zum Ende dieser Jahrgangsstufe soll das Niveau A2+ des GER (siehe Kapitel 3.1.2.) erreicht werden, weshalb es sinnvoll ist, bei der Suche nach Interviews im Korpus *Sacodeyl* im Bereich *Temas* nach *Rutinas y actividades cotidianas* oder *Hobbies* sowie im Bereich *Nivel CEF* nach Texten des Niveaus *A2* zu suchen. Die Schüler können sich in Paaren oder Kleingruppen unter den gefundenen Interviews einen Jugendlichen herausuchen, dessen Text sie sich genauer ansehen und später der Klasse präsentieren wollen.

Es besteht die Möglichkeit, dass die Schüler selbst im Korpus nach den Interviews suchen oder dass die Lehrkraft passende Interviews ausgedruckt mit in den Unterricht bringt, wobei die erste Option die Schüler mehr motiviert, vor allem, wenn sie das Interview ansehen und gleichzeitig mitlesen können. Wie bereits angedeutet, hat das Team des Korpus *Sacodeyl* leider Probleme mit dem Videoformat, wodurch die Videos derzeit nicht abgespielt werden können. Die hier vorgestellte Unterrichtseinheit zum Leben spanischer Jugendlicher kann zwar ebenso gut anhand der bloßen Texte durchgeführt werden, doch es ist für die Schüler viel interessanter, den jeweiligen Jugendlichen im Interview zu sehen und den Text am Computer gleichzeitig zu hören und mitzulesen. Es bleibt zu hoffen, dass die Videos in absehbarer Zeit ohne Probleme abgespielt werden können.

Außerdem könnten die muttersprachlichen Texte des Lernerkorpus CEDEL2 (siehe Kapitel 2.2.2.6.) bei landeskundlichen Themen im Unterricht verwendet werden. Die dort aufgelisteten Aufsätze sind zu allgemeinen Themen wie der Heimatregion, den Zukunftsplänen oder dem letzten Urlaub, aber auch zu drängenden Problemen und Fragen wie den Rechten homosexueller Paare oder der Legalisierung von Marihuana verfasst und können somit durchaus im Unterricht genutzt werden. Anhand dieser Texte können die Erfahrung oder auch die Meinung spanischer Muttersprachler zu diesen Themen in den Unterricht eingebunden und mit den eigenen Ansichten verglichen werden.

4.5.2. Umgangssprache und Jugendsprache

Im Lehrplan für Spanisch als spätbeginnende Fremdsprache heißt es in der 12. Jahrgangsstufe, also dem 3. Lernjahr, dass die Schüler auch Gespräche und Äußerungen von Muttersprachlern in regionalen Varianten oder der Jugendsprache verstehen sollen.¹²³⁷ In diesem Sinne könnte das Korpus COLA (siehe Kapitel 2.1.2.1.) für die Erarbeitung einiger Ausdrücke der Jugendsprache herangezogen werden. Die Schüler begegnen in diesen Gesprächen allerdings auch vielen Ausdrücken der Umgangssprache, die nicht speziell der Jugendsprache angehören, wodurch die Unterrichtseinheit allgemein zur Ausdrucksfähigkeit in gesprochenem Spanisch beiträgt.

Die Lehrkraft könnte den Schülern einige Gesprächsausschnitte zeigen oder diese selbst im Korpus suchen lassen. Die Schüler sollen die Texte durchsehen und unbekannte Ausdrücke erarbeiten.

¹²³⁷ Vgl. Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Spanisch (spätbeginnend) – Jahrgangsstufe 12*. Op. cit. (06.09.2017).

- T L 10 ASPEAKERS: <Tres chicos suben a un autobús/>
T L 10 MAESB2G03: ((Comment desc="9m 16.209 sg"/)) qué tal Física/
T L 10 MAESB2G02: 1[eh eh eh eh bueno yo he empezado]
MAESB2G01: 1[yo no he empezado]
T L 10 MAESB2G01: pero tengo aquí unos guiones de puta madre
T L 10 MAESB2G02: yo he empezado me he hecho yooo los guionees
T L 10 otro guión de campo eléctrico solo
T L 10 ah y momentos momentooooos no tengo ni puta idea
T L 10 MAESB2G01: yo tengo todo el examen resumido en cinco hojas

Abbildung 122: COLA – Ausschnitt aus einem Gespräch zwischen drei Jugendlichen über einen bevorstehenden Physiktest.¹²³⁸

In obigem Ausschnitt finden die Schüler beispielsweise zwei unterschiedliche Verwendungen des umgangssprachlich häufig gebrauchten *puta*. Während der erste, feststehende Ausdruck *de puta madre* die Bedeutung ‚toll‘, ‚cool‘ oder ‚super‘ trägt, bedeutet *ni puta idea* hingegen ‚keinen blassen Schimmer haben‘. Da Ausdrücke mit *puta* in der spanischen Jugendsprache wie auch generell in der Umgangssprache sehr geläufig sind, könnten sie den Schülern durchaus schon einmal begegnet sein.

- T L 10 MABPE2G01: me pone @nombre soy @nombre sí voy a bajar al metro luego
T L 10 MABPE2J01: luego . luego cuándo\ . éste @nombre es gilipollas
T L 10 <pausa/>
T L 10 ay Señor
T L 10 MABPE2G01: lo estás escuchando/
T L 10 MABPE2J01: no
T L 10 MABPE2G01: para por donde tu casa/ ... vamos @nombre
T L 10 <pausa/>
T L 10 joé tronca ves/ es que . siempre me pasa igual tronca
T L 10 .. qué asco chica .. y ha tenido que ser en tu casa
T L 10 MABPE2J01: . . sí te lo has tenido que dejar en mi casa

Abbildung 123: COLA – Ausschnitt aus einem Gespräch zwischen einem Jungen und einem Mädchen auf der Straße.¹²³⁹

¹²³⁸ Jørgensen, Annette Myre: *El corpus COLA. MAESB2-01b*. Online Publikation: <http://clu.uni.no/humfak/cola/txt-lyd/maesb2-01b.htm> (17.04.2018).

¹²³⁹ Jørgensen, Annette Myre: *El corpus COLA. MABPE2-11b*. Online Publikation: <http://clu.uni.no/humfak/cola/txt-lyd/mabpe2-11b.htm> (17.04.2018).

Anhand dieses zweiten Ausschnitts werden die Schüler auf Ausdrücke wie *gilipollas* ‚Vollidiot‘ und *tronca* ‚Freund‘ aufmerksam, die den deutschen Schülern weitaus weniger geläufig sein dürften. Bei *gilipollas* handelt es sich zwar um einen Ausdruck der generell in der Umgangssprache verwendet wird, doch *tronca* ist ein klarer Vertreter der Jugendsprache, der den Schülern vermutlich zum ersten Mal begegnet. Leider wird die Lesbarkeit der Transkripte aufgrund der Anonymisierung der vorkommenden Namen durch den Ausdruck *@nombre* beeinträchtigt.

Die Schüler sollen sich einige Texte durchlesen, Passagen daraus anhören und sich mit dem Inhalt auseinandersetzen. Unbekanntes Vokabular kann anhand eines Online-Wörterbuchs ermittelt werden. Die Ausdrücke der Jugendsprache sind teilweise ebenfalls in Wörterbüchern vermerkt, wenn nicht, so sollen die Schüler im Internet nach der Bedeutung suchen. Dort gibt es einige Seiten zur spanischen Jugendsprache, anhand derer sie die Bedeutung der Lexeme ermitteln können. Beispielsweise findet sich im Internet ein *Diccionario de la jerga juvenil*¹²⁴⁰, welches von einer Mittelschulklasse aus Navarra unter der Leitung ihrer Lehrerin erstellt wurde, sowie ein Eintrag über *Jerga juvenil* in *Wikilengua del español*.¹²⁴¹

Die Lehrkraft kann zuvor mit der Klasse vereinbaren, wie viele Ausdrücke jeder Schüler beziehungsweise jedes Schülerpaar ermitteln soll. Am Ende dieser oder in der nächsten Unterrichtsstunde sollen die gefundenen Ausdrücke anhand des jeweiligen Textausschnitts der Klasse vorgestellt und mittels der im Internet gefundenen Bedeutung erklärt werden. Das Forschen im Internet nach unbekannten Ausdrücken der Jugendsprache motiviert die Schüler dabei zunehmend, da sie keine vom Lehrer zuvor genau ausgefeilten Aufgaben lösen, sondern sich selbst auf die Suche begeben. Durch die Präsentation der Ausdrücke vor der Klasse beschäftigen sie sich noch eingehender mit dem Thema und sie können sich die Formulierungen besser merken. Zudem könnten die Schüler den jeweiligen Textausschnitt der Klasse szenisch präsentieren, da sie anhand der Konsequenzen wissen, in welcher Situation die Jugendlichen den Text äußern. Zur weitergehenden Beschäftigung mit dem umgangs- und jugendsprachlichen Wortschatz

¹²⁴⁰ Vgl. González Goñi, Silvia: *Diccionario de la jerga juvenil*. Online Publikation: https://issuu.com/euterpeee/docs/diccionario_jerga (25.02.2018).

¹²⁴¹ Vgl. Wikilengua del español (ed.): *Jerga juvenil / España*. Online Publikation: http://www.wikilengua.org/index.php/Jerga_juvenil/Espana (25.02.2018).

könnte die Lehrkraft die Schüler in Paaren oder Kleingruppen ein Gespräch erarbeiten lassen, in dem mindestens zehn der neuen Formulierungen verwendet werden sollen.

Die Tatsache, dass sie Gespräche lesen und hören, die von spanischen Jugendlichen in informellen Situationen in der Schule oder einem Jugendclub geführt wurden, dürfte für die Schüler aufgrund der Nähe zu ihrer eigenen Situation und ihren eigenen Gesprächen mit Gleichaltrigen sehr interessant sein. Neben der Erarbeitung informeller Ausdrücke, erfahren die Schüler auch Näheres über das Leben spanischsprachiger Jugendlicher und die Themen, welche diese beschäftigen. Dadurch ist es ihnen möglich, deren Leben mit ihrem eigenen vergleichen.

Allerdings muss hinzugefügt werden, dass die Audiodateien des COLA teilweise schwer zu verstehen sind wegen der Hintergrundgeräusche sowie der Tatsache, dass der außersprachliche Kontext fehlt. Aufgrund der Komplexität der Gespräche sollten nur fortgeschrittene Schüler ab dem Ende des 3. Lernjahrs selbstständig im Korpus suchen. Ansonsten kann die Lehrkraft passende Texte oder Passagen daraus auswählen und diese mit den Schülern zusammen im Unterricht besprechen.

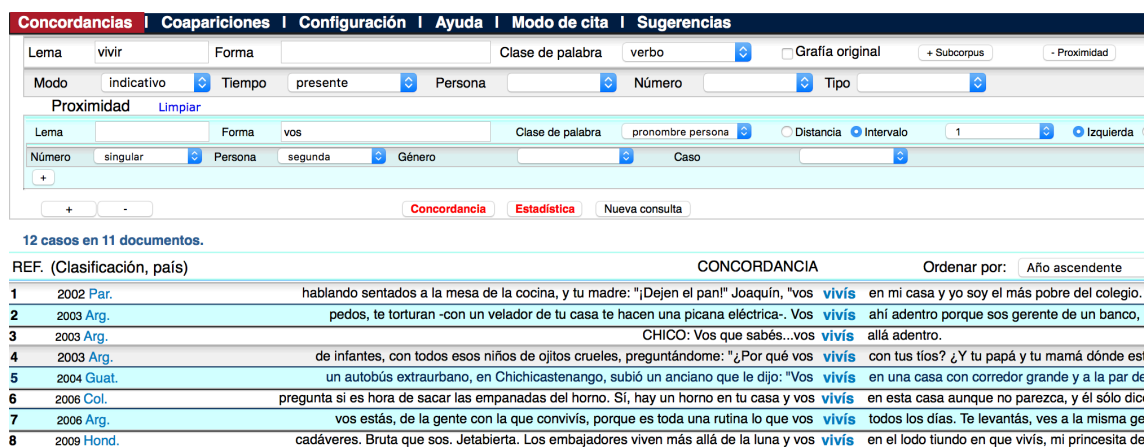
4.5.3. Argentinisches Spanisch

Bei der Bearbeitung des landeskundlichen Themas Argentinien ist es je nach Altersstufe und Lernstand der Schüler ab dem zweiten oder dritten Lernjahr sinnvoll, auch die Besonderheiten des argentinischen Spanisch anzusprechen. Ein Beispiel hierfür ist der *voseo*, bei dem unter anderem die 2. Person Singular nicht wie in der Standardsprache gebildet wird (*tú vives*), sondern eine vollkommen andere Verbform mit einem eigenen Personalpronomen verwendet wird (*vos vivís*) [*vos* = altspanisch, *vivís* = 2. Person Plural]. Außerdem ist das argentinische Spanisch durch einige Aussprachebesonderheiten gekennzeichnet. Bei der Bearbeitung dieser sprachlichen Unterschiede kann durchaus ein Korpus, wahlweise das CORPES (siehe Kapitel 2.1.1.2.), genutzt werden. So kann die Lehrkraft den Spanischlernern das Arbeitsblatt 23 – *El Español de Argentina* (Anhang 32) an die Hand geben und diese die Formen des *voseo* sowie die Ausspracheunterschiede selbst erarbeiten lassen. Generell wird der *voseo* im Unterricht lediglich thematisiert, sofern Formen desselben in einem Lehrbuchtext vorkommen, beispielsweise in Dialogen mit Argentinern. Die Lehrkraft weist in diesem Fall auf die Formen im Text hin und erklärt kurz die Besonderheit. Die Bearbeitung des

4. Nutzung von Korpora in der fremdsprachlichen Unterrichtspraxis

Arbeitsblatts 23 kann an die Behandlung eines solchen Texts anschließen, wobei die Schüler in diesem Fall von einer weitergehenden Einübung des *voseo* profitieren.

Für die Bearbeitung des Themas *voseo* soll ein Raster mit spanischen und argentinischen Formen ausgefüllt werden. Die spanischen Formen sind nicht angegeben, sondern sollen selbst eingetragen werden, um diese zu wiederholen. Sind sich die Schüler nicht sicher, können die Formen allerdings auch im Korpus abgerufen werden, auf dieselbe Weise wie die Formen des *voseo* im Folgenden ermittelt werden. Hierfür sollen die Schüler zunächst das Verb als Lemma in die Suchmaske eingeben und die Kategorien Wortart (*verbo*), Modus (*indicativo*) und Tempus (*presente*) auswählen. Nach einem Klick auf *Proximidad* erweitert sich die Suchmaske, das Personalpronomen *vos* wird als Form eingetragen und als 2. Person Singular gekennzeichnet (mit den sich eröffnenden Kategorien der *clase de palabra: número* und *persona*). Da auf dem Arbeitsblatt ebenfalls ein Screenshot wie der folgende angegeben ist, wissen die Schüler auf einen Blick, welche Auswahl sie treffen müssen:



Concordancias | Coapariciones | Configuración | Ayuda | Modo de cita | Sugerencias

Lema: Forma: Clase de palabra: ☐ Grafía original

Modo: Tiempo: Persona: Número: Tipo:

Proximidad

Lema: Forma: Clase de palabra: ☐ Distancia ☒ Intervalo ☒ Izquierda

Número: Persona: Género: Caso:

12 casos en 11 documentos.

REF. (Clasificación, país)	CONCORDANCIA	Ordenar por:	Año ascendente
1 2002 Par.	hablando sentados a la mesa de la cocina, y tu madre: "¡Dejen el pan!" Joaquín, "vos vivís en mi casa y yo soy el más pobre del colegio.		
2 2003 Arg.	pedos, te torturan -con un velador de tu casa te hacen una picana eléctrica-. Vos vivís ahí adentro porque sos gerente de un banco,		
3 2003 Arg.	CHICO: Vos que sabés...vos vivís allá adentro.		
4 2003 Arg.	de infantes, con todos esos niños de ojitos crueles, preguntándome: "¿Por qué vos vivís con tus tíos? ¿Y tu papá y tu mamá dónde es?		
5 2004 Guat.	un autobús extraurbano, en Chichicastenango, subió un anciano que le dijo: "Vos vivís en una casa con corredor grande y a la par de		
6 2006 Col.	pregunta si es hora de sacar las empanadas del horno. Sí, hay un horno en tu casa y vos vivís en esta casa aunque no parezca, y él sólo dijo		
7 2006 Arg.	vos estás, de la gente con la que convivís, porque es toda una rutina lo que vos vivís todos los días. Te levantás, ves a la misma ge		
8 2009 Hond.	cadáveres. Bruta que sos. Jetabierta. Los embajadores viven más allá de la luna y vos vivís en el lodo tiundo en que vivís, mi princesita de		

Abbildung 124: CORPES – Suche nach der Verbform der 2. Person Singular von *vibrar* im *voseo*.¹²⁴²

Sollten sich die Schüler bei einer der spanischen Verbformen nicht sicher sein, muss lediglich das Pronomen *vos* im Korpus durch das Pronomen *tú* ersetzt werden. Durch die Bearbeitung dieser Übung lernen die Schüler selbstständig die argentinischen Formen kennen, wodurch das Verständnis beim Lesen und Hören von Texten in argentinischem Spanisch erleichtert wird.

Im zweiten Teil des Arbeitsblatts sollen die Schüler Aussprachebesonderheiten des argentinischen Spanisch ermitteln. Hierfür wird das Subkorpus *Oral* des CORPES

¹²⁴² Real Academia Española (ed.): *CORPES. Entrada Compleja*. Op. cit. (31.08.2018).

genutzt. Da den fortgeschrittenen Spanischlernern im zweiten oder dritten Lernjahr die Lautschrift bereits aus anderen Sprachen, wie dem Englischen oder Französischen, geläufig ist, sollten die Lautschriftzeichen auf Arbeitsblatt 23 für sie kein Problem darstellen. Andernfalls sollte eine kurze Wiederholung genügen, um ihr Wissen aufzufrischen.

Für die Ermittlung der Ausspracheunterschiede sollen die Schüler das jeweilige Lexem als Lemma in die Suchmaske eintragen, dieses aber nicht im gesamten Korpus, sondern lediglich im mündlichen Subkorpus suchen. Daraufhin erscheinen die Konkordanzen aus unterschiedlichen Ländern, aus denen solche aus Spanien und Argentinien ausgewählt werden. Mittels eines Klicks auf einen der blau unterlegten Treffer gelangt man zur Transkription des Textabschnitts, wo man sich die Audiodatei anhören kann. Anhand des Vergleichs spanischer und argentinischer Beispiele soll somit die Besonderheit des argentinischen Spanisch ermittelt werden.

Im CORPES könnte auch direkt nach spanischen und argentinischen Beispielen gesucht werden, wie die folgende Abbildung zeigt, doch die vielen Einstellungen würden die Schüler zunehmend verwirren:

The screenshot shows the CORPES search interface. It features two identical search filter sections. The top section is for the lemma 'mayo' in the 'Forma' (Form) section. The bottom section is for the lemma 'mayo' in the 'Subcorpus' section. Both sections include a 'Limpiar' (Clear) button and a 'Subcorpus' dropdown menu. The filters are organized into several rows: 'Origen' (Origin) with options like 'América', 'España', 'Filipinas'; 'Zonas lingüísticas' (Linguistic zones) with options like 'Caribe continental', 'Chilena', 'México y Centroamérica', 'Río de la Plata'; 'Países' (Countries) with options like 'Argentina', 'Paraguay', 'Uruguay'; 'Medio' (Medium) with options like 'Escrito', 'Oral'; 'Tipología' (Typology) with options like 'Debate', 'Discurso', 'Entrevista'; 'Sexo' (Gender) with options like 'Hombre', 'Mujer', 'No indicado'; 'Grupo de edad' (Age group) with options like '0-14', '15-19', '20-34'; and 'Marcas' (Marks) with options like 'Cursiva', 'Negrita', 'Subrayado', 'Versalitas'. Below the filters, a table of search results is shown. The first five results are highlighted in blue and show the word 'mayo' in context. The results are ordered by 'Año ascendente' (Ascending year).

REF. (Clasificación, país)	CONCORDANCIA	Ordenar por: Año ascendente
21 2003 Arg.	fresco / la máxima no va a superar los veintiséis grados este lunes veintiséis de mayo	
22 2003 Arg.	este lunes / veintiséis de mayo	
23 2003 Arg.	en cuenta lo siguiente / no era la plaza de // de aquellas / este // eeh viejas plazas de Mayo que estaban de bote a bote	
24 2003 Arg.	experiencia única / poder salir al balcón / con un presidente // y ver la Plaza // eh de Mayo con un número importante de gente / es verdadera	
25 2003 Esp.	sábados / mañana tarde / mu muchos festivos a lo largo del año / en zonas turísticas / de mayo a septiembre / y por tanto / mmm esa búsqueda de	

Abbildung 125: CORPES – Suche nach *mayo* in gesprochenen Texten aus Spanien und Lateinamerika.¹²⁴³

¹²⁴³ Real Academia Española (ed.): *CORPES. Entrada Compleja*. Op. cit. (31.08.2018).

Eine derartige Übung vertieft nicht nur das Thema Landeskunde hinsichtlich der unterschiedlichen sprachlichen Entwicklung, sondern kann auch als Vorbereitung auf ein Hörverstehen in argentinischem Spanisch dienen. Die Besonderheiten des Spanischen Argentiniens können zwar zeitsparender von der Lehrkraft erklärt werden als die gezeigte Vorgehensweise mit einem Korpus, doch hier geht es erneut um eine Alternative, die es den Schülern ermöglicht, selbst die Unterschiede herauszufinden. Generell ist davon auszugehen, dass all das, was sich ein Lerner selbst erarbeitet hat, besser behalten werden kann als das, was er lediglich gehört hat. Anhand dieses Arbeitsblatts lernen die Schüler auch, dass sie mithilfe der Audiodateien des CORPES selbstständig herausfinden können, wie ein Lexem ausgesprochen wird, sofern sie sich diesbezüglich nicht sicher sind.

4.5.4. Mehrsprachigkeit in Spanien

Zudem ist es möglich, anhand von Korpora die unterschiedlichen regionalen Amtssprachen Spaniens von den Schülern selbst erforschen zu lassen, anstatt den Schülern nur die Unterschiede zu erklären und darauf hinzuweisen, dass ein Spanier, der ausschließlich des Kastilischen mächtig ist, Sprecher aus Katalonien, Galizien und dem Baskenland nur bedingt oder nicht versteht, sofern diese in ihrer regionalen Amtssprache kommunizieren. Die Lehrkraft könnte die Schüler selbst im spanisch-galicisch-katalanisch-baskischen Parallelkorpus CONSUMER, welches ein Subkorpus des CLUVI (siehe Kapitel 2.3.2.) mit mehr als 5,5 Mio. Lexemen ist, nach einem geeigneten und interessanten Textausschnitt suchen und diesen mit den Übersetzungen in den regionalen Amtssprachen vergleichen lassen. Sie sollen je nach Lernstand versuchen, die grundlegenden oder für sie auffälligsten Beziehungen zwischen den kastilischen Lexemen und deren galicischen, katalanischen sowie baskischen Entsprechungen herauszuarbeiten, und sie später gegebenenfalls der Klasse präsentieren.

Eine andere Möglichkeit ist eine gelenktere Vorgehensweise mithilfe eines Arbeitsblatts. Anhand des Arbeitsblatts 24 – *Los idiomas de España* (Anhang 33) sollen die Lerner im eben erwähnten Parallelkorpus CONSUMER nach Lexemen und Sätzen in den einzelnen Amtssprachen suchen. In der ersten Aufgabe geht es darum, sich mit dem Korpus und den unterschiedlichen Sprachen vertraut zu machen, indem beliebige Lexeme gesucht werden. Hierfür geben die Schüler Ausdrücke ihres Interesses in die spanische Zeile der Eingabemaske ein und vergleichen die ermittelten Ergebnisse im Hinblick auf die Unterschiede in den einzelnen Sprachen. In der neuen Version des CONSUMER

werden die Entsprechungen des Suchbegriffs in den anderen Sprachen in der Regel farbig markiert. Diese Markierung hilft dem Nutzer ungemein dabei, die jeweiligen Pendants zu ermitteln, doch in manchen Texten sind keine farbigen Hervorhebungen zu finden, weshalb man selbst die jeweilige Entsprechung suchen muss. Dabei fällt sehr schnell auf, dass die gesuchten spanischen Begriffe im Galicischen und Katalanischen problemlos ausfindig gemacht werden können, die Identifizierung der baskischen Entsprechungen allerdings Schwierigkeiten bereitet aufgrund der vollkommen unterschiedlichen Struktur der Sprache. Es geht hier jedoch nicht um die Ermittlung der einen korrekten baskischen Entsprechung eines kastilischen Lexems (welche mangels Kenntnis der baskischen Sprache sowohl Schülern als auch Lehrern große Probleme bereiten würde), sondern um die Erkenntnis, dass es sich hierbei um eine vollkommen andere Sprache handelt, bei der eben nicht einzelne Lexeme mit denjenigen der spanischen Sprache verglichen und die Bedeutung auf diese Weise hergeleitet werden kann. Aus diesem Grund ist die Aufgabe sehr frei gehalten und die Schüler sollen selbst die spanischen Lexeme auswählen, welche sie im Korpus suchen wollen. Zudem sind im Lösungsvorschlag verschiedene Formen für die baskischen Entsprechungen angegeben, um zu verdeutlichen, dass sich die Endung je nach dem Kontext, in welchem das Lexem erscheint, verändert. Am Ende der ersten Aufgabe sollen die Schüler die eben beschriebenen Erkenntnisse anhand der Beantwortung von drei kurzen Fragen festhalten.

In der zweiten Aufgabe sollen die Schüler vier Sätze der entsprechenden Sprache oder Varietät zuordnen. Es handelt sich dabei um den gleichen Satz, der eben nur in den vier unterschiedlichen Sprachen angegeben ist. Zum einen können die Schüler dies anhand ihres Wissens aus der ersten Teilaufgabe versuchen, zum anderen können sie den ganzen Satz oder auch Teile davon im Subkorpus CONSUMER suchen. Die Sätze der dritten Aufgabe sind ebenfalls entweder in der galicischen, der baskischen oder der katalanischen Sprache geschrieben, doch hierbei handelt es sich um unterschiedliche Sätze, deren Sprache sowie spanische Entsprechung die Schüler ermitteln sollen. Hier können die Schüler ebenfalls sowohl ihr Wissen aus den vorherigen Aufgaben als auch das Korpus für die Lösung heranziehen. Das Arbeitsblatt 24 kann ab dem Ende des zweiten beziehungsweise zu Beginn des dritten Lernjahrs im Unterricht verwendet werden.

Es handelt sich bei diesem Arbeitsblatt, wie auch beim vorherigen aus dem Kapitel 4.5.3., um eine zusätzliche Möglichkeit der Auflockerung. Die Besonderheiten und Unterschiede der regionalen Amtssprachen Spaniens könnte die Lehrkraft den Schülern in einem Lehrervortrag in kürzerer Zeit erklären und anhand von ein paar Beispielen veranschaulichen. Anhand dieses Arbeitsblatts soll lediglich ein Weg aufgezeigt werden, wie die Schüler selbst die Unterschiede ermitteln können, damit diese besser im Gedächtnis bleiben. Alternativ kann die Lehrkraft Lexeme, Phrasen oder ganze Textpassagen im Korpus suchen und sie den Schülern im Unterricht zeigen, ohne diese selbst im Korpus CLUVI suchen zu lassen. Einfach zu erschließende Texte, Textausschnitte oder auch nur einzelne Lexeme aus dem Korpus kann die Lehrkraft bereits Schülern des ersten Lernjahrs zeigen und ihnen damit die Unterschiede zwischen den einzelnen regionalen Amtssprachen nahebringen, wie dies im Lehrplan gefordert wird (siehe Kapitel 3.1.3.).¹²⁴⁴

Das *Corpus Trilingüe Paralelo Grial* (siehe Kapitel 2.3.1.) ist zwar auch ein Parallelkorpus, es umfasst jedoch nur die Sprachen Spanisch, Katalanisch und Englisch und besteht lediglich aus 2,3 Mio. Lexemen, welche zudem auf Texte aus technischen Ausbildungsrichtungen begrenzt sind. Aus diesem Grund kann es für Unterrichtseinheiten zu den regionalen Amtssprachen Spaniens nur im Bereich des Katalanischen und dessen Verhältnis zum Kastilischen verwendet werden.

In diesem Kapitel konnte bisher gesehen werden, wie die verfügbaren Korpora der spanischen Sprache für den Spanischunterricht genutzt werden können und was dabei beachtet werden sollte. Es gibt verschiedenste Möglichkeiten, wie ein Korpus für den Unterricht oder die Vorbereitung desselben verwendet werden kann. Dabei ist es nicht immer erforderlich, dass die Schüler selbst im Korpus forschen, doch es konnte anhand der Vorschläge in diesem Kapitel gesehen werden, dass Korpora nicht zu komplex sind, um sie zusammen mit Schülern zu konsultieren. Schüler können Korpora auf verschiedene Arten nutzen. Manche sind bereits zu Beginn des zweiten Lernjahrs möglich, andere erst in der Oberstufe, doch jede Nutzung der Korpora durch Schüler sollte im Idealfall dazu führen, dass die Schüler das Korpus auch für den eigenen

¹²⁴⁴ Vgl. Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Spanisch (3. Fremdsprache) – Jahrgangsstufe 8*. Op. cit. (06.09.2017).

Lernprozess außerhalb des Unterrichts verwenden. So können sie es zur selbstständigen Korrektur ihrer Textproduktionen, zur Überprüfung der Existenz bestimmter Konstruktionen sowie für Anregungen für die eigene Sprachproduktion heranziehen.

4.6. Eignung der bisher verfügbaren spanischen Korpora für den Unterricht

Im Folgenden soll ein Fazit zu den Korpora der spanischen Sprache gegeben werden, um in einer abschließenden Übersicht zu sehen, welche Korpora für den Unterricht geeignet sind und wofür sie genutzt werden können. Bei der Kategorisierung der Korpora wird dabei in der Regel nach der Auflistung im zweiten Kapitel vorgegangen. Jedes Korpus gehört allerdings immer verschiedenen Gruppierungen an, wie bereits in jenem Kapitel zu sehen war. Aus diesem Grund können einzelne Korpora bisweilen in anderen oder auch mehreren Kategorien erwähnt werden.

4.6.1. Referenzkorpora: CREA, CORPES und CdE:Orig

Generell sind Referenzkorpora am erfolgversprechendsten im Hinblick auf die Nutzung für den Fremdsprachenunterricht, da sie die spanische Sprache in ihrer Gesamtheit unter Beachtung der Repräsentativität abbilden, mitunter die umfangreichsten Korpora sind und zudem eine große Anzahl an Nutzungsmöglichkeiten bieten. Aufgrund ihrer Repräsentativität kann davon ausgegangen werden, dass die gefundenen Ergebnisse nicht nur für das Korpus zutreffen, sondern auch das Verhältnis in der gesamten spanischen Sprache widerspiegeln. Im Folgenden wird genauer auf die Vor- und Nachteile der Referenzkorpora CREA (siehe Kapitel 2.1.1.1.), CORPES (siehe Kapitel 2.1.1.2.) und den modernen Teil des CdE:Orig (siehe Kapitel 2.1.1.3.) eingegangen werden.

Das CORPES sowie beide Versionen des CREA bieten vielfältige Suchoptionen. So kann nach einem speziellen Titel, Autor oder Jahreszahlen gesucht sowie die Ergebnisse nach der Herkunft der Texte, dem Medium und dem Thema eingegrenzt werden. Das CORPES bietet zudem eine Suche nach der Textsorte. Im Gegensatz dazu ermöglicht das CdE:Orig lediglich die Einordnung nach Jahrhunderten sowie in die folgenden vier Bereiche: *oral, ficción, periódicos, académico*.

Die beiden annotierten Korpora der *Real Academia Española*, das CREA und das CORPES, sind vergleichbar hinsichtlich der Benutzeroberfläche und der Durchführung der Suchanfragen, doch das CORPES enthält die aktuelleren Texte. Das CORPES ist mit

derzeit ungefähr 286 Mio. Lexemen größer als das CREA mit 160 Mio. Lexemen. Zudem verfügt es als einziges Referenzkorpus über die Möglichkeit, sich Textausschnitte aus mündlichen Quellen anhören zu können. Der moderne Teil des CdE:Orig enthält lediglich 20 Mio. Lexeme, ist hinsichtlich der Benutzeroberfläche und der Durchführung der Suchanfragen anders gestaltet als die beiden Korpora der *Real Academia Española* und verfügt nicht über mündliche Texte, die sich der Nutzer anhören kann. Das CdE:Orig ist allerdings, vor allem in der Version 2008, sehr benutzerfreundlich und teilweise intuitiv zu bedienen. Die Auswahlmöglichkeiten in der neuen Version des CdE:Orig erübrigen die Verwendung regulärer Ausdrücke und erleichtern die Formulierung der Suchanfragen enorm. Es verfügt über zahlreiche Funktionen und ist übersichtlich gestaltet.

In Bezug auf das CREA sind derzeit noch beide Versionen online verfügbar. Man könnte meinen, dass es sinnvoller wäre, lediglich die neuere, annotierte Version des CREA zu verwenden, allerdings hat das CREA (Version 2015) auch seine Vorzüge. So ist es vor allem für die Arbeit mit Schülern geeignet, die noch nicht viel Erfahrung mit Korpora haben, da die Benutzeroberfläche sehr einfach gestaltet ist. Während die annotierte Version des CREA zahlreiche manuelle Einstellungen zulässt, mithilfe derer man die Suche genau eingrenzen kann und auch die Abstände zwischen zwei gesuchten Lexemen explizit angegeben werden können, besticht das alte CREA durch seine Schlichtheit. Dadurch, dass nur eine Eingabezeile vorhanden ist, können ganze Sequenzen von Lexemen eingegeben und gefunden werden. Durch diese einfache Vorgehensweise können die Schüler relativ schnell alleine mit dem Korpus umgehen. Zudem führt es Suchanfragen schneller aus als das neue CREA und das CORPES.

Bei CREA (Version 2015) sind allerdings die Trefferzahlen begrenzt, weshalb nur Ergebnisse mit weniger als 2000 Dokumenten oder 1000 Beispielen angezeigt werden können. Aus diesem Grund kann auch nicht unbegrenzt mit Platzhaltern gesucht werden, da die Suche nach Prä- und Suffixen zu viele Ergebnisse liefern würde. Dies ist das wohl größte Problem der alten Version des CREA, denn bei sehr frequenten Lexemen oder Phrasen können die Treffer nicht angezeigt werden aufgrund der Begrenzung der Anzeige der Konkordanzen. In diesen Fällen kann man aber dennoch zu Ergebnissen gelangen, indem die Suchanfrage genauer eingeschränkt oder ein Filter verwendet wird.

Da das CREA (Version 2015) außerdem nicht nach Wortarten getaggt oder lemmatisiert ist, kann nicht nach der relativen Häufigkeit eines Lemmas gesucht werden,

sondern nur nach sämtlichen vorkommenden Verbformen. Aufgrund der fehlenden Möglichkeit der Suche nach Wortarten, ist es ebenso wenig möglich, ambige Formen anhand der Wortart zu unterscheiden.

In der annotierten Version des CREA sowie im CORPES gibt es einige neue Optionen zur genaueren Suche, beispielsweise die Angaben zum Genus, zur Art des Substantivs (*propio* oder *común*) oder zu Tempus, Modus und Person bei Verben. Anhand dieser Möglichkeiten kann die Suche zunehmend verfeinert werden, um genau diejenigen Ergebnisse zu erhalten, die für die eigenen Untersuchungen relevant sind. Allerdings funktionieren diese Optionen der gezielten Suche noch nicht optimal, da aufgrund von Fehlern in der automatischen Annotation und Lemmatisierung unter den Ergebnissen häufig auch solche Konkordanzen erscheinen, die aufgrund der Sucheinstellungen ausgeschlossen sein sollten. Im CdE:Orig stehen zwar ebenfalls einige Auswahloptionen zur Verfügung, doch diese sind nicht ganz so detailreich wie diejenigen der annotierten Korpora der *Real Academia Española*. So können im CdE:Orig auch Genus und Numerus bei Adjektiven und Substantiven oder die Zeit bei Verben ausgewählt werden, doch bei Letzteren kann beispielsweise nicht nach der Person unterschieden werden. Die Formen der einzelnen Personen werden bei der Suche nach einer bestimmten Zeitform zunächst nach der Frequenz aufgelistet angezeigt und können dann ausgewählt werden. Anhand dieser Liste ist jedoch nicht ersichtlich, welche Form welcher Person entspricht.

Im CREA und im CORPES kann zudem mittels der Option *Proximidad* nach ganzen Folgen von Lexemen gesucht werden, indem der gewünschte Abstand zwischen den Lexemen eingegeben und bei der Suche berücksichtigt wird. In dieser Option kann auch allgemein nach Wortarten gesucht werden, beispielsweise nach Adjektiven in der Umgebung eines bestimmten Substantivs. Die detaillierte Eingabemaske ist ein Vorteil, gleichzeitig aber auch ein Nachteil. Durch diese Möglichkeit können Suchanfragen genauer gestellt werden, zugleich ist es aber nicht mehr möglich, ganze Phrasen in die Eingabezeile einzugeben, was die Suche komplexer macht und zugleich anfälliger für Fehler, die dazu führen können, dass keine Ergebnisse ermittelt werden. Das CdE:Orig hingegen lässt sowohl die Suche nach einzelnen Lexemen als auch nach Phrasen in einer einzigen Eingabezeile zu und ermöglicht die genauere Spezifizierung durch zusätzliche Optionen. Hier können somit ganze Phrasen eingegeben werden, während in den

4. Nutzung von Korpora in der fremdsprachlichen Unterrichtspraxis

annotierten Korpora der *Real Academia Española* jedes Lexem einzeln in einer separaten Eingabezeile der Option *Proximidad* eingetragen werden muss.

Zudem kann in den annotierten Korpora der *Real Academia Española* mittels der Optionen Y, O und NO gleichzeitig nach Beispielen gesucht werden, die beide Kriterien erfüllen (Y), nur eines der beiden (O) oder nur das erste, aber nicht das zweite Kriterium (NO). Das CdE:Orig erlaubt die Suche mit logischen Operatoren nicht, doch hier ermöglicht die Option *Comparar* einen Vergleich der Kollokate zweier Suchbegriffe.

Eine Option, die das CdE:Orig gegenüber den anderen Referenzkorpora auszeichnet, ist die Möglichkeit, semantisch-orientierte Suchanfragen durchzuführen. Hierbei handelt es sich um eine einzigartige und sehr hilfreiche Funktion mithilfe derer Synonyme gesucht werden können. Der Nutzer kann sich dabei nicht nur die Häufigkeit und Verteilung aller Synonyme eines bestimmten Lexems anzeigen lassen, sondern eben auch alle Lexeme, die zusammen mit diesen Synonymen auftreten, beispielsweise Substantive, die mit den Synonymen von *sagrado* verwendet werden, oder Adjektive mit den Synonymen von *imagen*.

In allen Referenzkorpora kann zudem mit Platzhaltern gesucht werden, um beispielsweise Wortfamilien zu erarbeiten oder bestimmte Wortbildungsmuster zu ermitteln. Im CREA (Version 2015) kann es bei Suchen mit Platzhaltern allerdings dazu kommen, dass die Ergebnisse aufgrund zu hoher Trefferzahlen nicht angezeigt werden können.

Im CORPES, wie auch im CREA, können *Coapariciones* mittels der statistischen Verfahren MI, *T-Score* und *LL-Simple* (siehe Kapitel 1.4.8.) ermittelt werden. Das CdE:Orig hingegen stellt die Ergebnisse lediglich nach dem Wert der MI geordnet dar, wobei die Zuordnung sich hier nicht derart problematisch darstellt, da der Abstand zwischen den beiden Lexemen frei gewählt werden kann. Der Abstand zwischen der Basis und dem Kollokator kann im CdE:Orig mit bis zu neun Lexemen je Seite angegeben werden, wohingegen in den Korpora der *Real Academia Española* eine Spanne von fünf Lexemen links und rechts des Suchbegriffs fest eingestellt ist und nicht verändert werden kann. Im CORPES und im CREA wird gleichzeitig nach allen Wortarten gesucht, wobei die Treffer nach der Wortart sortiert werden können. Im CdE:Orig können mittels eines Asterisks alle im Umfeld erscheinenden Wortarten ermittelt werden, doch es ist ebenso die Suche nach einzelnen Wortarten möglich.

Wenn im CORPES oder im CREA der Cursor auf einen der Treffer gezogen wird, zeigt das Korpus eine linguistische Analyse des Lexems an sowie den jeweiligen Textausschnitt, in dem es vorkommt. Zudem besteht in beiden Korpora die Möglichkeit der kombinierten Suche nach Lemmata und Formen sowie die explizite Suche nach der ursprünglichen Schreibweise. Letztere ist beispielsweise hilfreich bei Lexemen, die mit und ohne Akzent oder mit Groß- und Kleinschreibung auftreten können. Die zusätzliche Angabe der Wortart macht es möglich, bei Suchen zwischen Homonymen zu unterscheiden. Außerdem kann in diesen Korpora nach Ausdrücken von bis zu fünf Lexemen Länge gesucht werden, was beispielsweise bei Redewendungen wie *darse cuenta* sehr hilfreich sein kann. Im CdE:Orig hingegen wird mittels der Verwendung von eckigen Klammern nach Lemmata gesucht. Zugleich kann die Suche durch die Angabe der Wortart eingegrenzt werden, doch eine gleichzeitige Suche nach Lemma und Form ist nicht möglich. Dadurch, dass die Treffer zunächst nach deren Frequenz aufgelistet erscheinen, kann hier die gesuchte Form ausgewählt werden, sodass in den Konkordanzen nur die gewünschten Ergebnisse erscheinen. Diese Übersicht über alle Treffer, aufgelistet nach deren Frequenz, ist beispielsweise bei der Suche nach Kollokationen in einem bestimmten Intervall sehr hilfreich. Sucht der Nutzer in den Korpora der *Real Academia Española* mittels der Option *Proximidad* beispielsweise nach Adjektiven in der direkten Umgebung eines bestimmten Substantivs, so werden einige Konkordanzen wahllos angeordnet aufgelistet und man kann nicht erkennen, welche Kombinationen frequent sind und welche nicht. Um diese Angaben zu ermitteln, müsste man die infrage kommenden Kombination jeweils separat eingeben und danach suchen. Im CdE:Orig erscheint bei der Suche nach Kollokatoren in einem Abstand von einem Lexem links und rechts der Basis zunächst die beschriebene Übersicht mit Frequenzangaben, anhand derer relevante Kombinationen ausgewählt werden können, und im nächsten Schritt kann sich der Nutzer die Kollokationen zu eben diesen Kombinationen anzeigen lassen.

Die Aufteilung der Benutzeroberfläche des CdE:Orig (Version 2008) in ein Eingabefeld, eine Ergebnisanzeige und einen Informationsbereich, welche sich jeweils in einem eigenen Feld befinden, ist äußerst benutzerfreundlich, da sie es dem Suchenden erlaubt, gleichzeitig Informationen in die Suchmaske einzugeben, sich das Ergebnis anzeigen zu lassen sowie Informationen und Hilfestellungen zu lesen, sofern es Probleme oder Fragen gibt. Es ist manchmal bei der Betrachtung der Ergebnisse von Vorteil, die

zuvor eingegebenen Daten vor sich zu haben und gegebenenfalls bei problematischen Fällen auf den Informationsseiten nachzulesen. Die Hinweise zu Nutzungsmöglichkeiten, welche das Korpus für jede Auswahloption bereitstellt, sind äußerst hilfreich und bieten dem Nutzer Anregungen für eigene Suchanfragen. In der neuen Version des CdE:Orig sind die verschiedenen Felder in unterschiedlichen Reitern angeordnet, was den Vorteil bietet, dass jedes Feld über mehr Platz verfügt und der Nutzer dennoch die Eingaben in der Suchmaske überprüfen und gegebenenfalls für die nächste Suche abändern kann. Er muss allerdings zwischen den Reitern hin und her springen, denn man kann die Eingabemaske und die Ergebnisse nicht gleichzeitig betrachten. Die neue Benutzeroberfläche des CdE:Orig mutet zudem zunächst etwas fremd an, wenn man sich bereits an die alte Version von 2008 gewöhnt hat, da nur wenige Auswahlmöglichkeiten von vornherein angezeigt werden. Die meisten Optionen eröffnen sich erst durch weitere Klicks auf farbig gekennzeichnete Links wie *Secciones*, *Virtual*, *Orden/Límite* oder *Opciones*. Bei der alten Benutzeroberfläche werden die Optionen von vornherein angezeigt, man muss allerdings zunächst auf das Fragezeichen am rechten Rand der jeweiligen Eingabezeile klicken, damit sich die Eingabefelder öffnen. Generell ist zu sagen, dass das alte Interface des CdE:Orig entgegen der Aussage des Erstellers, Mark Davies, benutzerfreundlicher und intuitiver zu bedienen ist als das neue. Sollte der Nutzer allerdings mit einem Smartphone oder einem Tablet arbeiten, kann er lediglich die neue Version nutzen, da diese auch auf mobilen Geräten läuft.

Ein Vorteil des CdE im Hinblick auf die Benutzeroberfläche ist die Tatsache, dass diese in beiden Versionen nicht nur auf Spanisch, sondern auch auf Englisch verfügbar ist, was die Nutzung vor allem für Lerner mit geringen Spanischkenntnissen erleichtert. Des Weiteren werden die Ergebnisse schneller angezeigt als bei den Korpora der *Real Academia Española*.

Das CdE:Orig bietet neben der Anzeige der Ergebnisse als Liste auch die Darstellung als Grafik sowie die Anordnung der Treffer als *Palabra clave en contexto* an. Bei dieser Option kann der Benutzer angeben, nach welcher Umgebung, also nach dem ersten, zweiten oder dritten Lexem links oder rechts des Suchbegriffs, er suchen möchte, und er kann sich die Ergebnisse in alphabetischer Reihenfolge anzeigen lassen. Dabei werden die einzelnen Wortarten in unterschiedlichen Farben markiert.

Ein eindeutiger Vorteil der Korpora der *Real Academia Española* ist die freie Verfügbarkeit. Der Nutzer muss sich nicht anmelden, sondern kann das Korpus unverbindlich nutzen und unbegrenzt Suchanfragen durchführen. Diese freie Verfügbarkeit ist im Fall des CdE:Orig nicht gegeben (siehe Kapitel 2.1.1.3.). In der neuen Version des CdE:Orig erscheint außerdem nach je zehn bis fünfzehn Suchanfragen eine Meldung, ob der Nutzer nicht zu einem *premium account* wechseln und durch die Zahlung eines gewissen Beitrags die Korpora unterstützen möchte. Je häufiger das Korpus genutzt wird, desto öfter und länger werden die Warnungen angezeigt, wobei währenddessen das Korpus nicht genutzt werden kann. Innerhalb des *premium accounts* würden die Meldungen nicht mehr erscheinen, doch in der alten Version des CdE:Orig gibt es derartige Meldungen ohnehin nicht. Das CdE:Orig (Version 2008) bietet somit zwei nicht zu unterschätzende Vorteile gegenüber der neuen Version: eine geringere Wartezeit zwischen den Suchanfragen mit einer schnelleren Anzeige der Ergebnisse sowie die Übersichtlichkeit und Benutzerfreundlichkeit. Hinsichtlich der Lemmatisierung und Annotation sind beide Versionen gleich und auch die Ergebnisanzeige ist ähnlich.

Sofern die Lehrkraft das CdE:Orig mit der ganzen Klasse im Computerraum nutzen möchte, muss sich jeder Schüler beziehungsweise je ein Schüler pro Arbeitsgruppe an einem Computer registrieren. Die Klasse kann zusammen 250 Suchanfragen durchführen, das heißt, ein Arbeitsblatt könnte mit dieser Anzahl je nach Klassenstärke in Zweier- oder Dreiergruppen bearbeitet werden. Darüber hinaus wäre eine sogenannte *academic license* nötig, mittels derer jeder Schüler 200 und die Lehrkraft 400 Suchanfragen pro Tag durchführen könnte. Die Gebühr beträgt für Klassen von 30 Schülern 200 Dollar, für 80 Schüler 400 Dollar und darüber hinaus 600 Dollar pro Jahr.¹²⁴⁵ Nachdem die Schüler mit der Funktionsweise dieses sehr benutzerfreundlichen und übersichtlichen Korpus vertraut gemacht und einige Suchen zusammen durchgeführt wurden, können sie das Korpus mit ihrem Account allerdings auch in der kostenlosen Version zu Hause nutzen und haben somit jeweils mehr Suchanfragen frei, ohne einen Beitrag zahlen zu müssen.

Die neuen Korpora des CdE bieten viele Möglichkeiten, doch ein Nachteil im Vergleich zur alten Version sind die eben erwähnten, relativ häufig erscheinenden Hinweise auf die finanzielle Unterstützung des Korpus, welche bei der Nutzung eines

¹²⁴⁵ Vgl. Davies: *Corpus.byu.edu. Academic license*. Op. cit. (26.01.2018).

premium accounts oder einer *academic license* entfallen. Bei einer Nutzung für schulische Zwecke kann jedoch nicht davon ausgegangen werden, dass Geld für die Nutzung eines Korpus ausgegeben wird. Für gelegentliche Suchen im Korpus können diese Hinweise akzeptiert werden, doch bei extensiver Nutzung des Korpus sollte entweder der Beitrag gezahlt oder ein anderes Korpus genutzt werden.

Die Lehrkraft sollte sich das Korpus zuvor genau ansehen, überprüfen, ob es den eigenen Anforderungen und den Möglichkeiten, wie man es mit der Klasse nutzen möchte, entspricht, doch aufgrund der vielen Optionen die das CdE bietet, wäre die *academic license* tatsächlich eine Überlegung wert, wenn man das Korpus öfter verwenden möchte. Die Zahlung des Beitrags ermöglicht außerdem die Nutzung der höheren Zahl an Suchanfragen in allen Korpora von Mark Davies, also nicht nur im CdE:Orig, sondern auch im CdE:Nuevo (siehe Kapitel 2.6.1.), im *Google Books: Spanish* (2.6.2.) und im *NOW Corpus: Spanish* (2.6.3.). Sofern das Korpus jedoch nur wenige Male für die Bearbeitung eines Arbeitsblatts genutzt werden soll, reicht der kostenlose Zugang aus.

Jedes der Referenzkorpora weist Vor- und Nachteile auf. Da die Nutzungsmöglichkeiten des CORPES diejenigen des CREA übersteigen und es sich zudem um das neuere der beiden Referenzkorpora der *Real Academia Española* handelt, ist es sinnvoll mit einer Kombination aus dem CORPES und dem CdE:Orig zu arbeiten. Mittels dieser beiden Korpora können viele Besonderheiten der spanischen Sprache ermittelt sowie umfangreiches Unterrichtsmaterial erstellt werden. Diese beiden Korpora eignen sich sowohl als Referenzquelle für die Lehrkraft als auch für die Nutzung durch Schüler. Großes Bedauern muss allerdings darüber geäußert werden, dass in keinem der Referenzkorpora die gesamten Texte zugänglich sind, sondern lediglich Textabschnitte mit einer Länge von wenigen Sätzen angezeigt werden können.

Bei der Verwendung der Referenzkorpora durch die Schüler selbst stellt sich allerdings das dort anzutreffende Niveau der spanischen Sprache aufgrund der vermehrten Aufnahme schriftlicher Textgattungen für jüngere Schüler als problematisch heraus, da im Unterricht zumindest in den ersten Jahren vor allem die gesprochene Sprache vermittelt wird. Wie in den Aussagen der Redaktion Spanisch des Cornelsen Verlags in Kapitel 3.3.2.2.1. bereits angemerkt wurde, sind die in diesen Korpora vermittelten Sprachauschnitte sowohl hinsichtlich des Niveaus als auch in Bezug auf die

Themen teilweise zu komplex für Schüler im zweiten oder auch dritten Lernjahr. Daher sollte die Arbeit mit diesen Korpora von der Lehrkraft mit Bedacht ausgewählt und didaktisch sinnvoll eingesetzt werden.

Es bleibt zu hoffen, dass in naher Zukunft ein Korpus wie das französische CRFC (siehe Kapitel 3.3.2.2.1.), bei dem ein ebenso großer Anteil an geschriebenen wie an gesprochenen Texten zu finden ist und letztere derart viele Transkripte von Fernsehsendungen umfassen, für das Spanische erstellt wird. Da das CRFC mit Augenmerk auf die Entwicklung von korpusbasierten Nachschlagewerken und Lehrbüchern, die Verbesserung von Unterrichtsmaterialien sowie für die Unterrichtsplanung und als Korrekturhilfe erstellt wurde, ist zu erwarten, dass es den Französischunterricht in jeder Hinsicht voranbringen wird. Ein solches im Hinblick auf den Fremdsprachenunterricht erstelltes Referenzkorpus wäre für die spanische Sprache ebenso wünschenswert und könnte wie das CRFC von Sprachwissenschaftlern, Lexikographen, Lehrwerkerstellern, Lehrern und Schülern genutzt werden.

4.6.2. Spezialkorpora: COLA und *Spanish in Texas Corpus*

Bei der Nutzung von Spezialkorpora im Unterricht müssen die Lehrkraft und die Schüler immer vor Augen haben, dass es sich bei der darin dargestellten Sprache, anders als bei Referenzkorpora, nicht um einen repräsentativen Ausschnitt der spanischen Sprache, sondern um eine besondere Varietät, einen Dialekt oder einen sich anderweitig von der Standardvarietät abgrenzenden Sprachausschnitt handelt. Daher können anhand dieser Korpora keine Aussagen über die Sprache im Allgemeinen getroffen, sondern es können lediglich Besonderheiten der jeweiligen Varietät erarbeitet werden.

Im Fall des COLA (siehe Kapitel 2.1.2.1.) handelt es sich um die Darstellung der Sprache spanischer Jugendlicher. Die Dateien des Korpus COLA sind aufgrund der Tatsache, dass es sich um gesprochene Texte der Jugendsprache handelt, sehr interessant für deutsche Jugendliche. Sie können unter anderem nach den Informanten und deren Alter ausgewählt werden und die Suchmaske erlaubt eine gleichzeitige Suche nach bis zu drei Lexemen.

Das COLA könnte für die Durchführung eines Hörverstehens verwendet werden, insbesondere für allgemeine Hörverstehensaufträge, wie Fragen nach dem Thema der Konversation oder der Meinung der Sprechenden, da es vollständige Audiodateien und deren Transkripte zur Verfügung stellt. Außerdem können die Dateien des COLA dazu

genutzt werden, um allgemein die Aussprache und Intonation der Lerner zu verbessern, da die Schüler durch das Anhören der Texte nicht nur die Aussprache, sondern beispielsweise auch die Satzmelodie der spanischen Jugendlichen hören und verinnerlichen. Es könnte zudem aufgrund der Möglichkeit, nach Lexemen zu suchen, die daraufhin in Konkordanzen inklusive Audiodateien angezeigt werden, für die Schulung der Aussprache spezieller Lexeme genutzt werden. Es ist jedoch nicht so umfangreich wie das CORPES, weshalb manche Lexeme nicht enthalten sein könnten.

Besonders geeignet scheint es für die Erarbeitung von Ausdrücken der Jugend- und Umgangssprache, weshalb die Nutzung allgemein zur Ausdrucksfähigkeit im gesprochenen Spanisch beitragen kann. Ebenso könnten anhand des Korpus COLA Diskursmarker, Füllwörter und typische Formulierungen für Ausrufe oder Einwürfe erarbeitet werden, da es sich um reale, spontane Gesprächssituationen handelt.

Während sich das COLA mit der Jugendsprache beschäftigt, präsentiert das multimodale *Spanish in Texas Corpus* (siehe Kapitel 2.1.2.2.) spanische und bilingual spanisch-englische Sprachausschnitte aus Interviews und Gesprächen zwischen Sprechern aus unterschiedlichen Regionen in Texas. In diesem Korpus gibt es keine Suchmaske, wodurch die Texte auch nicht nach Themen, Sprechern oder Lexemen durchsucht werden können. Die Dateien bieten allerdings neben Audiodateien und Transkripten auch Videomitschnitte der Interviews, sodass anhand dieses Korpus Hörverstehensaufgaben durchgeführt werden können, sofern es sich nicht um die Standardvarietät handeln muss. Zudem ist es möglich, anhand der Texte Inhalte zum Leben spanischer und bilingual spanisch-englischer Texaner zu erarbeiten. Die Dateien sind jedoch mit einer Dauer von neun Minuten bis zu einer Stunde zu lang für ein Hörverstehen im Unterricht und müssten daher gekürzt werden.

Aufgrund der Tatsache, dass in diesem Korpus nicht nach einzelnen Lexemen gesucht werden kann, ist es nicht möglich, die Aussprache einzelner Lexeme anhand des Korpus zu erarbeiten. Im Allgemeinen können die Texte aber zur Nachahmung der Aussprache und Intonation genutzt werden, sofern die Lehrkraft für das jeweilige Unterrichtsthema den Sprachausschnitt und die bilinguale Herkunft der Sprecher für geeignet hält.

4.6.3. Opportunistische Korpora: *esTenTen*, *Spa_news_2011_3M*, *Spa_newscrawl_2011*, *Hemero*, *CEA*, *CdE:Nuevo*, *Google Books: Spanish* und *NOW Corpus: Spanish*

Auch opportunistische Korpora bieten keine Repräsentativität im Hinblick auf die Gesamtheit der spanischen Sprache, da das Hauptaugenmerk bei ihrer Erstellung auf der Größe liegt. Dabei sind manche von ihnen, wie das *CdE:Nuevo* (siehe Kapitel 2.6.1.) oder das *Hemero* (siehe Kapitel 2.4.1.2.), durchaus für den Unterricht zu nutzen, doch andere, wie das *Spa_news_2011_3M* (siehe Kapitel 2.1.3.1.), sind nur bedingt für den schulischen Alltag geeignet. Zudem gibt es Korpora, die für den Unterricht passend wären, jedoch kostenpflichtig sind.

Zu letzteren zählt das Korpus *esTenTen* der Online-Plattform *SketchEngine* (siehe Kapitel 2.8.1.1.). Das Korpus bietet unter anderem die Option, sich eine Auflistung der signifikantesten Kollokationen geordnet nach der Wortart des Kollokators anzeigen zu lassen. Des Weiteren kann das *esTenTen* als Thesaurus genutzt werden, da es in der gleichnamigen Funktion eine Liste von ähnlichen Lexemen anzeigt. Aufgrund der vielen sich in diesem Korpus bietenden Möglichkeiten, wäre es für den Spanischunterricht sehr interessant, doch die Beitragskosten machen eine Nutzung im schulischen Kontext schwer möglich.

Das *Spa_news_2011_3M* enthält zwar eine große Menge an Texten, aufgrund der mangelnden Annotation und der Tatsache, dass es nur aus Zeitungstexten aus dem Jahr 2011 besteht, von denen man sich noch dazu lediglich einzelne Sätze anzeigen lassen kann, ist dieses Korpus allerdings nur bedingt für den Fremdsprachenunterricht nutzbar. Sollte sich der Nutzer lediglich nach der Verwendungsweise bestimmter Lexeme erkundigen oder deren Existenz nachweisen wollen, kann das Korpus jedoch durchaus verwendet werden. Außerdem ist das Korpus sehr interessant für Nutzer, deren Augenmerk auf dem Gebrauch des Spanischen in Zeitungstexten liegt.

Mit dem *Spa_newscrawl_2011* (Wortschatz Universität Leipzig, siehe Kapitel 2.1.3.2.) verhält es sich ähnlich. Hier handelt es sich ebenfalls um Zeitungstexte aus dem Jahr 2011 und auch hier sind keine vollständigen Texte zugänglich. Doch während das *Spa_news_2011_3M* immerhin lemmatisiert ist und dort nach ganzen Phrasen gesucht werden kann, ist hier lediglich die Suche nach einzelnen Lexemen möglich, denn das *Spa_newscrawl_2011* ist weder lemmatisiert noch in irgendeiner Weise annotiert. Es

enthält jedoch neben den Konkordanzen Informationen zum sprachlichen Kontext des Suchbegriffs sowie einen Graphen mit Lexemen, die in dessen Umfeld häufig erscheinen.

Das *Hemero* hingegen ist anders aufgebaut als die beiden zuvor beschriebenen opportunistischen Zeitungskorpora. Es ist mit 660 Mio. Lexemen um ein Vielfaches größer als das *Spa_news_2011_3M* (3 Mio.) und das *Spa_newscrawl_2011* (17 Mio.) und dort sind die vollständigen Texte sowie Links zu den Originalzeitungsartikeln verfügbar. Anhand dieses Korpus können Artikel aus spanischen, argentinischen und mexikanischen Zeitungen gesucht werden. Auf diese Weise erhält der Nutzer mit nur einer Suchanfrage Informationen zu Artikeln aus mehreren spanischsprachigen Zeitungen. Außerdem kann mit dem eingegebenen Suchbegriff nach dem Inhalt des Texts oder nach dem Titel gesucht werden. Das Korpus ist allerdings weder lemmatisiert noch auf irgendeine Weise annotiert.

Das Korpus *Hemero* ist sehr gut geeignet als Textquelle für Aufgaben zum Leseverstehen oder für landeskundliche Inhalte, doch es enthält lediglich Texte aus den Jahren 1997 bis 2009. Zudem wird auf der Internetseite angegeben, dass es weiterhin ausgebaut wird¹²⁴⁶, doch diese Erweiterung war zumindest für die Zeit von Juli 2017 bis Juli 2019 nicht zu erkennen. Ein derartiges Korpus aus aktuellen Texten, sozusagen ein Monitorkorpus aus Zeitungstexten, wäre für den Unterricht äußerst wünschenswert.

Im Gegensatz zu den vier bereits beschriebenen Korpora, bei denen es sich zum einen um ein zwar annotiertes, jedoch kostenpflichtiges Korpus und zum anderen um drei mehr oder weniger umfassende Zeitungskorpora ohne spezielle Annotation handelt, sind die vier folgenden Korpora sowohl kostenlos verfügbar als auch lemmatisiert und morphosyntaktisch annotiert. Das CEA (siehe Kapitel 2.4.1.1.) ist mit 540 Mio. Lexemen ein sehr umfangreiches Korpus des geschriebenen, zeitgenössischen Spanisch, das sich allerdings aus sehr komplexen Texten wie Akten und Gesetzestexten aus dem europäischen Parlament zusammensetzt. Es ermöglicht Suchen nach Lexemen, Lemmata, Wortarten und Konstruktionen unter Verwendung regulärer Ausdrücke. Zudem können Frequenzlisten und Kollokationen abgerufen werden.

Es muss allerdings angemerkt werden, dass das CEA nicht intuitiv zu bedienen und damit nicht so benutzerfreundlich ist wie das CdE oder die Korpora der *Real Academia Española*, da die Eingabesprache komplexer ist, wie in Kapitel 4.2.3. zu sehen war. Wenn

¹²⁴⁶ Vgl. Molino de Ideas (ed.): *Hemero.es - El corpus molinero*. Op. cit. (12.07.2019).

der falsche Modus angeklickt ist oder ein Zeichen des Suchausdrucks falsch beziehungsweise eine Leerstelle zu viel oder zu wenig angegeben ist, liefert die Suche keine Ergebnisse. Aus diesem Grund ist in der vorliegenden Arbeit auf Arbeitsblätter verzichtet worden, welche die Schüler anhand des CEA bearbeiten sollen. Für Lehrer ist es ein umfangreiches Korpus, welches durchaus verwendet werden kann, doch für die Nutzung mit Schülern birgt es zu viele Fehlerquellen und ist zu komplex gestaltet. Des Weiteren ist es aufgrund der Textzusammensetzung thematisch ungeeignet für den Schulunterricht, und es ist teilweise sehr langsam in der Ermittlung der Ergebnisse.

Das CdE:Nuevo zeichnet sich im Vergleich zu den bisher in diesem Kapitel beschriebenen opportunistischen Korpora durch seine Aktualität, aber auch durch seine Größe aus, denn es umfasst 2 Milliarden Lexeme aus Internettextrn der Jahre 2013 und 2014. Da das CdE:Nuevo ein opportunistisches Korpus aus aktuellen Internetseiten ist, handelt es sich bei den angezeigten Textpassagen um aktuelle Informationen zum betreffenden Thema. Während es für manche Bereiche, wie beispielsweise Kollokationen, über die man eine Auskunft im Hinblick auf deren tatsächliches Vorkommen in der Gesamtheit der Standardvarietät erhalten möchte, nur bedingt geeignet ist, da es sich eben um ein opportunistisches und kein repräsentatives Korpus handelt, ist dies genau der Vorteil, wenn die Schüler selbst im Korpus suchen und anhand der Konkordanzen grammatikalische oder lexikalische Inhalte erschließen sollen, denn es finden sich aktuelle Informationen im Korpus. Da es sich dabei zudem um Sprachauschnitte aus Internettextrn handelt und nicht um literarische Texte, sind die Konkordanzen für Schüler besser verständlich.

Aufgrund seiner Aktualität und Größe, der geographischen Vielfalt und der zahlreichen Nutzungsmöglichkeiten ist es auf jeden Fall ein für den Nutzer sehr wertvolles Korpus und sollte durchaus für den Spanischunterricht verwendet werden, auch wenn es nicht als Referenzkorpus eingeordnet werden kann. Im Gegensatz zu anderen Korpora ermöglicht das CdE:Nuevo gegen eine Gebühr von 245-795 \$ auch die Einsicht in die vollständigen Texte.¹²⁴⁷ Diese Kosten sind für die schulische Nutzung jedoch zu hoch, da das Korpus an sich auch ohne die Gebühr genutzt werden kann. Generell verhält es sich beim CdE:Nuevo ebenso wie bei den anderen Korpora von Mark Davies, denn es kann nur nach einer Registrierung genutzt werden, wobei ein

¹²⁴⁷ Vgl. Davies: *Full-text corpus data from corpus.byu.edu – Purchase data*. Op. cit. (13.12.2016).

4. Nutzung von Korpora in der fremdsprachlichen Unterrichtspraxis

angemeldeter Nutzer, der keinen Beitrag zahlt, eine gewisse Anzahl an Suchanfragen frei hat und darüberhinaus den Beitrag eines *premium account* oder einer *academic license* zahlen müsste (siehe Kapitel 4.6.1.).

Das CdE:Nuevo bietet zudem das Programm *WordAndPhrase*, welches es ermöglicht, ganze Texte hochzuladen und mit den Informationen aus dem Korpus abzugleichen. Mithilfe dieser Möglichkeit, einen Text in das Programm zu laden, sich die Häufigkeit der einzelnen Lexeme des Texts sowie deren Frequenz im Korpus anzeigen zu lassen, Einträge einzelner Lexeme durchzusehen und nach Synonymen im CdE:Nuevo suchen zu können, kann die Lehrkraft unbekannte oder sich wiederholende Lexeme eines Texts, aber auch ganze Phrasen, durch Synonyme ersetzen oder die gefundenen Synonyme für Vokabelangaben nutzen. Aufgrund der Benutzerfreundlichkeit und Übersichtlichkeit könnte das Korpus beziehungsweise das Programm *WordAndPhrase* auch fortgeschrittenen Schülern vorgestellt werden, damit diese es für die Verbesserung ihrer eigenen Sprachproduktionen nutzen können.

Das *Google Books: Spanish* (siehe Kapitel 2.6.2.), ein weiteres Korpus von Mark Davies, ist mit 45 Milliarden Lexemen das größte in dieser Übersicht aufgeführte Korpus, doch es unterscheidet sich auch grundlegend von den anderen Korpora, da es sich um keine Textsammlung handelt, in der systematisch nach bestimmten Lexemen oder Ausdrücken gesucht werden kann, sondern um eine Art Suchmaschine, welche die Einträge von *Google Books* durchsucht und als Ergebnis eine Liste mit Links zu eben diesen Einträgen anzeigt. Es ist geeignet für die Ermittlung von einzelnen Lexemen oder Kollokationen, welche als Liste oder Grafik angezeigt werden können, bietet aber nicht alle Suchoptionen der anderen Korpora von Mark Davies, wie in Kapitel 2.6.2. gesehen werden konnte.

Das neueste der in dieser Arbeit präsentierten Korpora ist das *NOW Corpus: Spanish* (siehe Kapitel 2.6.3.). Es handelt sich hierbei um ein sehr großes opportunistisches Korpus, welches darüber hinaus kontinuierlich jeden Monat um ungefähr 150 Mio. Lexeme erweitert werden soll. Es ist ähnlich zu nutzen wie das CdE:Nuevo, da es ebenfalls aus Internettexten besteht und über die gleiche Eingabemaske verfügt. Sollte das Korpus tatsächlich monatlich erweitert werden, ist es auch für den Sprachunterricht sehr interessant, da es zur Klärung von Fragen zum aktuellen Sprachstand oder zum Sprachwandel herangezogen werden kann. Zudem können die

Erstellung virtueller Korpora sowie der zeitliche und geographische Vergleich auch im Sprachunterricht genutzt werden. Das *NOW Corpus: Spanish* ist zwar nicht so groß wie das *Google Books: Spanish*, es ist im Gegensatz zu Letzterem allerdings ein richtiges Korpus mit allen Suchoptionen, die für gewöhnlich in Davies Korpora vorhanden sind, und nicht lediglich eine Suchmaschine für die Inhalte von *Google Books*.

Je nachdem, wofür das Korpus genutzt werden soll, können die opportunistischen Korpora also durchaus im und für den Unterricht verwendet werden, insbesondere das Korpus *Hemero* für die Erstellung von Aufgaben zum Leseverstehen. Lediglich zwei dieser Korpora können annähernd wie ein Referenzkorpus genutzt werden – das CdE:Nuevo und das *NOW Corpus: Spanish*. Sie bieten zahlreiche Funktionen, mithilfe derer im Korpus geforscht werden kann, und liefern sowohl interessante als auch aussagekräftige Ergebnisse, sofern der Nutzer bedenkt, dass es sich hierbei um den Bereich der spanischen Sprache in Online-Artikeln und nicht um die Standardvarietät handelt. Darüber hinaus ist das CdE:Nuevo durch das zusätzliche Programm *WordAndPhrase* auch für Schüler sehr interessant.

4.6.4. Historische Korpora: CORDE, CDH, CdE:Orig (Gesamt), *Biblia Medieval* und ADMYTE

Die im zweiten Kapitel aufgelisteten historischen Korpora sind teilweise mit ihren Pendanten im Bereich der Referenzkorpora vergleichbar. So ist das CORDE (siehe Kapitel 2.2.1.1.) von der Benutzeroberfläche her fast mit dem CREA (Version 2015, siehe Kapitel 2.1.1.1.) identisch und es hat zudem die gleichen Funktionen. Es ist weder lemmatisiert noch in irgendeiner Weise annotiert und die Trefferzahlen sind auf Ergebnisse mit weniger als 2000 Dokumenten oder 1000 Beispielen begrenzt. Aus diesem Grund können Suchen mit Platzhaltern oder nach Prä- oder Suffixen teilweise nicht angezeigt werden, da sie zu viele Ergebnisse liefern würden. In diesen Fällen muss ein Filter angewandt oder der Suchausdruck abgeändert werden.

Das CDH (siehe Kapitel 2.2.1.2.) hingegen ist wie das CORPES (siehe Kapitel 2.1.1.2.) oder die annotierte Version des CREA gestaltet. Es ist lemmatisiert sowie morphosyntaktisch annotiert, bietet die gleichen Funktionen wie die beiden Referenzkorpora und ist somit ebenso für die Suche nach Kollokationen geeignet. Allerdings überschneidet es sich hinsichtlich der Texte mit CREA und CORDE.

4. Nutzung von Korpora in der fremdsprachlichen Unterrichtspraxis

Das CdE:Orig (Gesamt, siehe Kapitel 2.2.1.3.) ist die vollständige Fassung des Korpus CdE:Orig (siehe Kapitel 2.1.1.3.) und kann folglich wie der Teil des modernen Spanisch des CdE:Orig genutzt werden. Es bietet damit vielfältige Nutzungsmöglichkeiten, wie alle Korpora von Mark Davies, doch es ist erst nach Anmeldung und dann nur begrenzt kostenlos nutzbar.

Die beiden verbleibenden Korpora, *Biblia Medieval* (siehe Kapitel 2.2.1.4.) und ADMYTE (siehe Kapitel 2.8.1.2.), haben kein Pendant unter den Referenzkorpora, sondern es handelt sich bei ihnen ausschließlich um Spezialkorpora. Das im Internet kostenlos und ohne Anmeldung zugängliche Korpus *Biblia Medieval* ist als Parallelkorpus Spanisch-Latein-Hebräisch nutzbar, doch aufgrund der Thematik des Korpus und der Tatsache, dass es sich um Texte aus dem Mittelalter handelt, ist dieses Korpus kaum im Spanischunterricht zu verwenden.

Das Korpus ADMYTE ist für Studenten des Spanischen, vor allem für diejenigen, die ihren Schwerpunkt auf das Altspanische legen, sicherlich eine geeignete Textquelle. Aufgrund der Tatsache, dass die spanische Sprache des Mittelalters im Unterricht jedoch gar nicht behandelt wird und man für die Nutzung des Korpus eine relativ hohe Gebühr entrichten muss (siehe Kapitel 2.8.1.2.), ist dieses Korpus für den Sprachunterricht von geringem Nutzen.

Während historische Korpora für Sprachwissenschaftler eine unschätzbare Quelle für die Erforschung vergangener Sprachstadien darstellen und damit Aufschluss über die Entwicklung der spanischen Sprache geben, können sie im Fremdsprachenunterricht an Schulen kaum genutzt werden. Im Lehrplan wird lediglich für die 11. und 12. Jahrgangsstufe im Fach Spanisch als dritte Fremdsprache angegeben, es sollen im Unterricht mindestens zwei klassische Werke der spanischen Literatur und Geistesgeschichte anhand von Auszügen besprochen werden, beispielsweise Cervantes, Lope de Vega oder Calderón de la Barca.¹²⁴⁸ Diese Werke werden den Schüler allerdings in modernem Spanisch vorgelegt. Sofern nun die Lehrkraft den Schülern zeigen möchte, dass die spanische Sprache zu dieser Zeit nicht diejenige war, wie sie sie heute kennen, könnte sie der Klasse einen Ausschnitt aus einem Original vorlegen. Leider sind in den hier aufgeführten historischen Korpora jedoch keine vollständigen Texte zugänglich,

¹²⁴⁸ Vgl. Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Spanisch (3. Fremdsprache) – Jahrgangsstufe 11/12*. Op. cit. (06.09.2017).

sondern es können lediglich Konkordanzen angezeigt werden. Aus diesem Grund sind die historischen Korpora, obgleich sie für das Studium der spanischen Sprache viele Vorteile mit sich bringen mögen, für den Spanischunterricht an Schulen nicht geeignet. Sollte die Lehrkraft den Schülern die Entwicklung eines Lexems veranschaulichen oder mit ihnen das erste Auftreten bestimmter Lexeme erforschen wollen, so wird sie auf eines der großen Korpora CORDE, CDH oder CdE:Orig (Gesamt) zurückgreifen. Die historischen Spezialkorpora ADMYTE und *Biblia Medieval* sind für Studenten des Spanischen sehr empfehlenswert, für den Schulalltag jedoch zu speziell.

4.6.5. Lernerkorpora: CORELE, SPLLOC, LANGSNAP, *Fono.ele*, CAES, CEDEL2 und SPT

Wie in Kapitel 4.2.6. zu sehen war, können Lernerkorpora für die Erstellung von Arbeitsblättern zur Fehlerkorrektur genutzt werden. Neben der Verwendung für Unterrichtsmaterial kann dieser Korpustyp aber auch zum Training von Lehrkräften herangezogen werden, insbesondere im Hinblick auf die Fehlerkorrektur im Mündlichen. Bei den Korpora CORELE (siehe Kapitel 2.2.2.1.), SPLLOC (2.2.2.2.), LANGSNAP (2.2.2.4.) und SPT (2.4.3.3.) sind alle Interviews als Audiodatei und Transkript abrufbar. Von diesen Korpora enthält allerdings nur das CORELE Daten deutscher Spanischlerner und auch nur in diesem Korpus kann nach Formen, Lemmata oder Wortarten gesucht werden. Das *Fono.ele* (siehe Kapitel 2.2.2.3.) verfügt zwar über Audiodateien von Lernern mit deutscher Muttersprache, es sind jedoch keine Transkripte verfügbar, weshalb auch nicht nach Lemmata oder Wortarten gesucht werden kann. Im CAES (siehe Kapitel 2.2.2.5.) ist die Suche nach Lemmata und Wortformen möglich, doch die Treffer werden lediglich als Konkordanzen angezeigt, der Nutzer erhält keinen Zugang zu den vollständigen Texten. Zudem enthält es keine Daten deutscher Muttersprachler. Im CEDEL2 (siehe Kapitel 2.2.2.6.), welches wie das CAES ein Korpus der geschriebenen Sprache darstellt, sind weder Daten deutscher Muttersprachler enthalten, noch ist die Suche nach Lemmata oder Wortarten möglich. Für Arbeitsblätter zur Fehlerkorrektur kann derzeit daher lediglich auf das CORELE zurückgegriffen werden.

Generell ist für das Training der Fehlerkorrektur das SPT besonders gut geeignet, da es eigens zu diesem Zweck erstellt wurde. Auf der Internetseite des SPT kann der Nutzer seine Einschätzung zu Grammatik, Vokabular, Flüssigkeit und Gesprächsführung angeben und sich danach von der Webseite Rückmeldung geben lassen, wie akkurat er bewertet hat. Dieses Korpus wäre ideal für die Lehrerbildung, doch es enthält lediglich

Daten englischer Muttersprachler und kann daher, insbesondere im Hinblick auf die Aussprache, die Bedürfnisse deutscher Lehrkräfte nicht befriedigen. Für allgemeine Einschätzungen kann es jedoch sehr wohl genutzt werden. Es kann zudem von fortgeschrittenen Spanischlernern verwendet werden, um ihre Kenntnisse zu erweitern.

Leider ist das Material des CORELE in den einzelnen Muttersprachen nicht besonders umfangreich. Wie bei allen Sprachen in diesem Korpus handelt es sich in Bezug auf Lerner mit deutscher Muttersprache nur um vier Interviews. Das Korpus ist aber sehr gut aufgebaut und annotiert, weshalb es durchaus für den Unterricht genutzt werden kann. Neben der Erstellung von Arbeitsblättern zur Fehlerkorrektur kann die Lehrkraft den Schülern beispielsweise Ausschnitte von Interviewten mit unterschiedlicher Muttersprache vorspielen, sie diese mit denen der spanischen Kontrollgruppe vergleichen und daraus Rückschlüsse auf ihre eigene Aussprache ziehen lassen. Zudem können die Schüler häufige Fehler in den einzelnen Kategorien ausfindig machen und versuchen, diese zu korrigieren, bevor sie sich die Verbesserung anzeigen lassen. Es wäre allerdings wünschenswert, ein größeres Fehlerkorpus nach diesem Vorbild an der Hand zu haben.

Für Hörverstehensübungen ist das Korpus SPLLOC mit seinen realen, spontanen Gesprächssituationen gut geeignet, da es auch Daten spanischer Muttersprachler enthält. Es kann aus verschiedenen Geschichten, Interviews und Paardiskussionen ausgewählt werden und die Audiodateien, mit einer Dauer von zwei bis acht Minuten, lassen sich von der Länge her gut in den Fremdsprachenunterricht integrieren. Anhand des SPLLOC kann zudem die Aussprache geübt werden, wobei sich der Nutzer lediglich die vollständigen Texte anhören kann und nicht, wie im CORPES, einzelne Lexeme oder Phrasen. Die Transkripte des SPLLOC sind allerdings nicht einfach mitzulesen, wie in Kapitel 4.3. beschrieben.

Die Vorgehensweise des LANGSNAP, von Spanischlernern vor, während und nach ihrem Auslandsaufenthalt mündliche und schriftliche Texte zu erfassen und daran die Entwicklung des Gebrauchs der Fremdsprache sowie des Wissens über dieselbe zu dokumentieren, ist sehr interessant im Hinblick auf den Spracherwerb. Allerdings handelt es sich hierbei erneut lediglich um englische Muttersprachler, sodass für den Erwerb des Spanischen durch deutsche Muttersprachler keine Rückschlüsse gezogen werden können und das Korpus für deutsche Schüler nicht als Fehlerkorpus genutzt werden kann. Die

Dateien der zehn spanischen Muttersprachler, die zu Vergleichszwecken erhoben wurden, können jedoch für ein Hörverstehen oder zur Schulung der Aussprache genutzt werden, da sie hinsichtlich der Länge und der Thematik durchaus in den Unterricht integriert werden können. Die Transkripte können allerdings, ebenso wie im SPLLOC, nicht problemlos mitgelesen werden, wie in Kapitel 4.3. zu sehen war.

Das *Fono.ele* kann zwar von Lehrkräften zur Einschätzung des Sprachniveaus ihrer Schüler und zur Einübung der mündlichen Fehlerkorrektur genutzt werden, da es Daten von Spanischlernern mit deutscher Muttersprache umfasst, doch aufgrund des Mangels an Transkripten kann es weder zur Erstellung von Arbeitsblättern noch zur Durchführung eines Hörverstehens genutzt werden.

Im CEDEL2 werden die vorhandenen Texte der spanischen Muttersprachler und der Spanischlerner vollständig angezeigt und können zusammen mit allen relevanten Angaben zum jeweiligen Text heruntergeladen werden. Zudem sind die dort vorzufindenden Aufsätze zu interessanten Themen wie der Heimatregion, einer berühmten Persönlichkeit, einem vor kurzer Zeit gesehenen Film oder dem vergangenen Urlaub, aber auch zu drängenden Problemen der Gegenwart wie Drogenkonsum oder Terrorismus verfasst worden. Somit können die muttersprachlichen Texte durchaus in Bezug auf die Landeskunde für den Unterricht verwendet werden, insbesondere bei Themen wie der Heimatregion oder beim Vergleich unterschiedlicher Meinungen spanischer Muttersprachler zu Themen wie der Einwanderung.

Wie in diesem Unterkapitel zu sehen war, können die verfügbaren spanischen Lernerkorpora zu verschiedenen Zwecken im und für den Unterricht eingesetzt werden. Wie bereits in Kapitel 4.2.6. angemerkt, wäre allerdings die Verfügbarkeit eines größeren Lernerkorpus, besonders im Hinblick auf Lerner mit deutscher Muttersprache, wünschenswert.

4.6.6. Multilinguale Korpora: CLUVI und *Corpus Trilingüe Paralelo Grial*

Die beiden Parallelkorpora *Corpus Trilingüe Paralelo Grial* und CLUVI (siehe Kapitel 2.3.1. und 2.3.2.), deren Hauptaugenmerk auf den regionalen Amtssprachen Spaniens sowie deren Verhältnis zum Kastilischen und anderen europäischen Sprachen liegt, können aufgrund ihrer Spezialisierung auf die eben genannten Varietäten nicht so gut in den Schulalltag eingebettet werden wie große Referenzkorpora oder Lernerkorpora. Dennoch könnte zumindest das CLUVI, wie in Kapitel 4.5.4. zu sehen

war, in eine Unterrichtseinheit über die verschiedenen Regionen Spaniens und deren Nationalsprachen Eingang finden.

Das CLUVI beschäftigt sich mit dem Galicischen und dessen Verhältnis zum Spanischen, zu anderen europäischen Sprachen sowie zu den weiteren regionalen Amtssprachen Spaniens. Insbesondere das darin enthaltene Subkorpus CONSUMER kann für eine Einheit zum Vergleich der unterschiedlichen Amtssprachen Spaniens genutzt werden, wie in Kapitel 4.5.4. zu sehen war.

Das *Corpus Trilingüe Paralelo Grial* ist zwar auch ein Parallelkorpus, aufgrund der Tatsache, dass es nur spanische, katalanische und englische Texte umfasst, ist es für eine Unterrichtseinheit zu den regionalen Amtssprachen Spaniens allerdings nur bedingt nutzbar, wie in Kapitel 4.5.4. zu sehen war.

Leider gibt es derzeit kein spanisch-deutsches Parallelkorpus, welches für den Sprachunterricht genutzt werden könnte. Ein derartiges Übersetzungskorpus mit Texten auf Spanisch und Deutsch, der Muttersprache der Schüler, wäre äußerst wünschenswert und könnte im Unterricht zu verschiedensten Zwecken, beispielsweise für Aufgaben zur Mediation, genutzt werden.

4.6.7. Korpora gesprochener Sprache: *Macrocorpus*, C-ORAL-ROM, PRESEEA, CORLEC und C-Or-DiAL

Obwohl die Mehrheit der Korpora sich mit geschriebener Sprache beschäftigt, gibt es durchaus einige Korpora der gesprochenen spanischen Sprache, wie bereits im zweiten Kapitel zu sehen war, und auch unter den Referenzkorpora existiert ein Korpus, welches über mündliche Texte sowie dazugehörige Audiodateien verfügt, das CORPES (siehe Kapitel 2.1.1.2.).

Wie in diesem vierten Kapitel gesehen werden konnte, können Korpora der gesprochenen Sprache im Allgemeinen für Aufgaben zur Schulung des Hörverstehens oder auch zur Erarbeitung und Verbesserung der Aussprache genutzt werden. Für diese Zwecke können die Dateien der Muttersprachler in den Lernerkorpora SPLLOC (siehe Kapitel 2.2.2.2.) und LANGSNAP (2.2.2.4.) sowie das Korpus der Jugendsprache COLA (2.1.2.1.) verwendet werden, auf welche bereits in den Kapiteln 4.6.5. beziehungsweise 4.6.2. genauer eingegangen wurde.

Leider sind einige der mündlichen Korpora nur auf CD-ROM verfügbar und daher nicht für alle Lehrkräfte zugänglich (siehe Kapitel 4.6.10.). Dabei handelt es sich um das

Macrocorpus (siehe Kapitel 2.4.2.1.), das C-ORAL-ROM (2.4.2.2.) und das PRESEEA (2.8.2.1.). In Bezug auf das C-ORAL-ROM ist zudem zu sagen, dass die Verfügbarkeit eines Vergleichskorpus der gesprochenen romanischen Sprachen sehr bemerkenswert und für sprachwissenschaftliche Anliegen auf jeden Fall von höchstem Nutzen ist. Für schulische Zwecke ist die Nutzbarkeit aufgrund der begrenzten Zugänglichkeit auf CD-ROM jedoch sehr eingeschränkt. Zudem ist die Tatsache, dass es sich um ein Vergleichskorpus handelt, für den Unterricht nur zweitrangig, da es sich lediglich um gesprochene Texte romanischer Sprachen handelt und nicht um einen möglichen Vergleich zwischen dem Spanischen und dem Deutschen.

Die im Internet verfügbaren Audiodateien und Transkripte des Korpus PRESEEA sind von ihrer Länge her gesehen durchaus geeignet für ein Hörverstehen im Unterricht (siehe Kapitel 4.4.1.). Es ist allerdings sehr schwer, die Dateien thematisch einzuordnen und passende Texte für den Unterricht zu finden, da sie zwar nach verschiedensten Kriterien sortiert werden können (siehe Kapitel 2.8.2.1.), allerdings nicht nach dem Thema des gesamten Interviews. Eine Suche nach Schlagwörtern ist möglich, doch dabei handelt es sich nur um einzelne Begriffe, die im jeweiligen Text enthalten sind. Eine Auswahl kann die Lehrkraft also nur treffen, indem sie die Texte liest beziehungsweise anhört und dabei auf deren Nutzbarkeit für den jeweiligen Unterrichtsabschnitt achtet.

Das Korpus CORLEC (siehe Kapitel 2.4.2.3.) ist zwar online zugänglich, es kann allerdings nicht für ein Hörverstehen genutzt werden, da sich keine Audiodateien auf der Internetseite befinden. Die freie Verfügbarkeit aller Transkriptionen im Internet ist beachtenswert und die Texte können für sprachwissenschaftliche Zwecke durchaus gewinnbringend genutzt werden, doch die fehlende Möglichkeit der Suche im Korpus mangels einer Suchmaske macht das Korpus für schulische Zwecke uninteressant und schwer nutzbar.

Das Korpus C-Or-DiAL (siehe Kapitel 2.4.2.4.) hingegen könnte durchaus für die Erstellung eines Hörverstehens genutzt werden, sofern die Lehrkraft den Schülern spontane gesprochene Sprache und kommunikative Funktionen vermitteln möchte. Wie es auf der Internetseite heißt, sollen die spontanen Sprachaufnahmen die Mängel didaktischer Materialien ausgleichen sowie das Verständnis und die Ausdrucksweise der

Schüler fördern.¹²⁴⁹ Die Dateien sind ohne die gleichzeitige Lektüre des Transkripts allerdings schwer zu verstehen, aufgrund des Sprechtempos, der sich überschneidenden Sprecherwechsel und der Kommentare der anderen Gesprächsteilnehmer. Zudem sind die Transkripte aufgrund der Einteilung in Gesprächsabschnitte anstatt Sätzen nicht leicht mitzulesen. Die Lehrkraft sollte bei der Auswahl von Dateien aus diesem Korpus daher sehr bedacht vorgehen.

Unter den Korpora der gesprochenen Sprache sind derzeit somit die beiden Lernerkorpora SPLLOC und LANGSNAP, sowie das Korpus der Jugendsprache COLA am besten für die Nutzung im Unterricht geeignet. Ansonsten können für Aufgaben zum Hörverstehen auch die multimodalen Korpora aus dem nächsten Kapitel verwendet werden.

4.6.8. Multimodale Korpora: SPT, *Sacodeyl* und *Backbone Spanish*

Wie soeben angesprochen, können multimodale Korpora sehr gut für die Erstellung eines Hörverstehens genutzt werden. Zu ihnen zählen das *Sacodeyl*, das *Backbone Spanish* sowie das SPT (siehe Kapitel 2.4.3.). Während es sich bei den ersten beiden Korpora um Aufnahmen spanischer Muttersprachler handelt, ist das SPT ein reines Lernerkorpus ohne Aufnahmen von Muttersprachlern zum Abgleich. Aus diesem Grund können die Korpora nicht für dieselben Zwecke verwendet werden.

Das SPT ist speziell für das Training von Spanischlehrkräften erstellt worden. Das Korpus ist sehr übersichtlich aufgebaut und die Videos lassen sich problemlos abspielen, was bei den anderen multimodalen Korpora nicht der Fall ist. Dieses Korpus kann als äußerst gewinnbringend für die Lehrerausbildung bezeichnet werden, doch es umfasst lediglich Daten englischer Muttersprachler und kann damit nicht von deutschen Lehrkräften für die Verbesserung der Korrektur herangezogen werden.

Das Korpus *Sacodeyl* wäre ein ideales Korpus für den Fremdsprachenunterricht, sowohl für die spanische Sprache als auch für Englisch, Französisch und Italienisch. Aufgrund der für die Interviews ausgewählten Themen, die sich größtenteils mit dem Lehrplan decken (siehe Kapitel 3.1.3.), und der Option, die Videos sowohl nach dem Niveau des GER (siehe Kapitel 3.1.2.) als auch nach grammatikalischen und lexikalischen Themen auswählen zu können, wäre dieses Korpus eine gewinnbringende

¹²⁴⁹ Vgl. Laboratorio de Lingüística Italiana (ed.): *Corpus Didáctico Anotado Lingüísticamente (C-Or-DiAL)*. Online Publikation: <http://lablita.dit.unifi.it/corpora/cordial/> (19.03.2017).

Erweiterung für den Fremdsprachenunterricht. Leider lassen sich die Videos derzeit nicht abspielen, wie bereits mehrfach bemängelt. Wenn die Spanischlerner Interviews mit spanischen Jugendlichen anhören und gleichzeitig das Transkript lesen beziehungsweise sogar die jeweiligen Videos ansehen könnten, würden sie zum einen Informationen über das Leben der Jugendlichen in Spanien erhalten und zum anderen ihre rezeptiven Fähigkeiten trainieren sowie anhand des muttersprachlichen Vorbilds ihre Aussprache verbessern. Die Verfügbarkeit eines Hörtexts und des dazugehörigen Transkripts würde den Schülern das Verstehen enorm erleichtern und durch das Ansehen eines Videos würden die Aufgaben zudem attraktiver.

Bisher sind also lediglich die Texte des *Sacodeyl* zu nutzen. Es lässt zwar nur die Suche nach Formen und nicht nach Lemmata oder Wortarten zu, da es weder lemmatisiert noch morphosyntaktisch annotiert ist, doch die Konkordanzen können hinsichtlich der Satzlänge, der Komplexität und der Themen direkt für Arbeitsblätter übernommen werden, unter Umständen auch auf Anfängerniveau. Während Beispielsätze aus Referenzkorpora teilweise auf einem zu hohen sprachlichen Niveau und damit hinsichtlich des Wortschatzes sowie der verwendeten grammatischen Strukturen zu schwierig sind für die Schüler, verwenden die spanischen Jugendlichen in den Interviews einfache Satzmuster und gängiges Vokabular aufgrund der alltäglichen Themen, die auch deutschen Jugendlichen nicht unbekannt sind.

Sofern die Videos des *Sacodeyl* problemlos abgespielt werden könnten beziehungsweise sofern Audiodateien vorhanden wären, könnte dieses Korpus als unschätzbare Quelle für Hörverstehensübungen bezeichnet werden und wäre zudem der Vermittlung landeskundlicher Themen dienlich. Es wäre daher sehr wünschenswert, wenn die Ersteller die Videos so formatieren könnten, dass das Korpus in vollem Umfang genutzt werden kann. Auch eine Vergrößerung des Korpusumfangs oder die Erstellung eines weiteren Korpus mit ähnlicher Zielsetzung wäre für den Sprachunterricht sehr förderlich.

Im zweiten muttersprachlichen, multimodalen Korpus, dem *Backbone Spanish*, gibt es ähnliche Probleme mit dem Abspielen der Videos. Hier ist es zwar nicht der Fall, dass diese gar nicht angesehen werden können, aber sie werden häufig nur stockend wiedergegeben, sodass im Unterricht die ebenso auf der Seite verfügbaren herunterladbaren Audio- und Videodateien genutzt werden sollten.

Das *Backbone Spanish* kann sehr gut für Hörverstehensübungen genutzt werden, da neben den Audio- und Videodateien auch Transkripte ohne zusätzliche Angaben über die gesprochene Sprache verfügbar sind und die einzelnen Texte beziehungsweise Textabschnitte nach den Niveaustufen des GER eingeteilt sind. Zudem sind die Texte in Themen wie Kultur, Wirtschaft, Soziales und Bildung kategorisiert. Auch die Aussprache könnte anhand dieses Korpus geübt werden, sofern sich der Nutzer ganze Texte oder Ausschnitte daraus anhören möchte und nicht speziell nach der Aussprache einzelner Lexeme sucht. Die Texte des *Backbone Spanish* sind sogar nach diatopischen Varietäten eingeteilt. So gibt es neben dem neutralen Spanisch unter anderem auch Spanisch mit andalusischem oder galicischem Akzent.

Das *Sacodeyl* und das *Backbone Spanish* sind derzeit die einzigen pädagogischen Korpora, welche im Spanischunterricht genutzt werden können. Die Entwicklung weiterer Korpora mit Suchoptionen nach lehrplankonformen Themen und Kompetenzniveaus wäre für den Sprachunterricht von großem Vorteil.

4.6.9. Annotierte Korpora

Wie in den vorherigen Kapiteln zu sehen war, sind viele Korpora auf irgendeine Weise annotiert. Die spanischen Referenzkorpora sind heutzutage beispielsweise alle lemmatisiert und morphosyntaktisch annotiert. Im Folgenden werden insbesondere die annotierten Spezialkorpora verglichen und hinsichtlich ihrer Nutzungsmöglichkeiten für den Fremdsprachenunterricht beschrieben.

4.6.9.1. Morphosyntaktisch annotierte Korpora: CT-IULA und *El Grial*

Eine morphosyntaktische Annotation ist bei den meisten Korpora durchaus sinnvoll, da sie beispielsweise Suchen nach Wortarten ermöglicht. Meist fällt diese Art der Annotation mit der Lemmatisierung zusammen, sodass in den betreffenden Korpora nicht nur nach einzelnen Wortformen, sondern auch nach Lemmata gesucht werden kann. Während die ersten Korpora der *Real Academia Española*, das CREA (Version 2015, siehe Kapitel 2.1.1.1.) und das CORDE (siehe Kapitel 2.2.1.1.), zunächst nicht annotiert waren, sind die neue Version des CREA sowie das CORPES (siehe Kapitel 2.1.1.2.) und das CDH (siehe Kapitel 2.2.1.2.) lemmatisiert und morphosyntaktisch annotiert. Wie in diesem vierten Kapitel zu sehen war, sind Korpora mit morphosyntaktischer Annotation für die Nutzung im Fremdsprachenunterricht besonders geeignet, beispielsweise bei der

Erarbeitung von Kollokationen (siehe Kapitel 4.1.1.) oder der Ermittlung von Wortfamilien (siehe Kapitel 4.1.5.2.).

Zudem gibt es zwei Spezialkorpora, die morphosyntaktisch annotiert sind: das CT-IULA und *El Grial* (siehe Kapitel 2.5.1.). Beide Korpora könnten aufgrund der Möglichkeiten, die ihre Annotation bietet, ebenso gut für die Bearbeitung einiger der in diesem Kapitel vorgestellten Arbeitsblätter, beispielsweise des AB 14 zu den Präpositionen, genutzt werden, doch sie bieten nicht den Bedienkomfort des CdE:Nuevo, mit dem dieses Arbeitsblatt erstellt wurde und mit dem es bearbeitet werden sollte. Die Eingabe ist bei den Korpora CT-IULA und *El Grial* komplexer als im CdE:Nuevo und in den Referenzkorpora, wie in Kapitel 4.2.1. zu sehen war.

Da die großen Referenzkorpora den Vorteil bieten, dass sie repräsentativ gegenüber der spanischen Sprache sind, ist es sinnvoller, Arbeitsblätter oder Aufträge, für welche es eines morphosyntaktisch annotierten Korpus bedarf, anhand eines derselben zu bearbeiten anstatt sich in die Funktionsweise dieser ohnehin komplexeren Spezialkorpora einzuarbeiten. Zudem ist es möglich, die Referenzkorpora für verschiedene Ziele einzusetzen, weshalb eine Einarbeitung in diese Korpora sowohl für die Lehrkraft als auch für die Schüler sinnvoller ist. Letztere sollten nicht mit einer Fülle an Korpora, die für spezielle Zwecke geeignet sind, überhäuft werden, sondern es ist didaktisch und pädagogisch ratsam, im Unterricht mit wenigen Korpora, die mehrere Nutzungsmöglichkeiten bieten, zu arbeiten.

Lehrkräfte können durchaus auf die beiden Korpora CT-IULA und *El Grial* zurückgreifen. Sie ermöglichen eine Suche nach mehreren Variablen, wodurch nach Kombinationen verschiedener Wortarten, wie eben auf Verben folgende Präpositionen, gesucht werden kann. Im Korpus *El Grial* können Sequenzen von sechs Einheiten und im Fall des CT-IULA sogar beliebig viele Einheiten gesucht werden.

Das Korpus *El Grial* bietet sogar eine Funktion namens *El Manchador*, mit welcher der Nutzer gleichzeitig nach verschiedenen Sequenzen, die mehrere *tokens* umfassen können, suchen und sich diese farblich getrennt anzeigen lassen kann. Diese Funktion ist beispielsweise für den Vergleich ähnlicher Strukturen, wie *para* mit Infinitiv und *para que* mit *subjuntivo* geeignet, um infolgedessen Näheres über die Unterschiede und Anschlussmöglichkeiten zu erfahren. Für Schüler hätte die Einarbeitung in dieses Korpus allerdings einen geringen Nutzen.

4.6.9.2. Syntaktisch annotierte Korpora: BDS, *Europarl-es* und AnCora

Unter den syntaktisch annotierten Korpora sind BDS, *Europarl-es* sowie AnCora zu nennen (siehe Kapitel 2.5.2.). Generell ist anzumerken, dass im Unterricht nahezu keine syntaktischen Analysen durchgeführt werden, weshalb die Vorteile und Besonderheiten syntaktisch annotierter Korpora in der Schule kaum genutzt werden können.

Das Korpus BDS ist ein manuell annotiertes syntaktisches Spezialkorpus. In Bezug auf Größe (116.000 Sätze) und Suchmöglichkeiten ist es allerdings begrenzt und die jeweiligen Einstellungen würden die Schüler überfordern, da diese sich mit den syntaktischen Fachbegriffen im Unterricht für gewöhnlich nicht auseinandersetzen. Daher eignet es sich nicht für schulische Zwecke.

Das *Europarl-es* ist sowohl morphosyntaktisch als auch syntaktisch annotiert und mit seinen 27,2 Mio. Lexemen größer als das Korpus BDS. Es bietet zudem mehr Suchoptionen, da beispielsweise die Wortart, morphologische und syntaktische Informationen sowie Platzhalter in die Suche miteinbezogen werden können. Auch die Suche nach mehrgliedrigen Ausdrücken ist anhand mehrerer Eingabefenster mit detaillierten Einstellungsoptionen möglich sowie die Ermittlung von Sätzen, in denen Lexeme bestimmte syntaktische Funktionen übernehmen, zum Beispiel solche mit einem Infinitiv als Subjekt des Satzes. Allerdings kann auch dieses Korpus nicht zur Erreichung der Unterrichtsziele eingesetzt werden, da es zu komplex ist und die Schüler sich im Unterricht nicht mit syntaktischen Analysen auseinandersetzen. Lehrkräfte können dieses Korpus durchaus zur Überprüfung von Hypothesen verwenden, da es, anders als das Korpus BDS, vielfältige Suchoptionen bietet und auch eine beachtliche Größe hat.

Das Korpus AnCora ist nicht nur morphosyntaktisch und syntaktisch, sondern auch semantisch getaggt. Obwohl das Korpus auf unterschiedlichste Art annotiert ist, sind die Such- und Darstellungsoptionen sehr begrenzt, wie in Kapitel 2.5.2.3. zu sehen war. Die einzelnen Sätze werden allerdings bei der Suche nach bestimmten syntaktischen Funktionen als eine Art Baumdiagramm dargestellt. Das Korpus ist für Studenten der Hispanistik sicherlich eine interessante Informationsquelle, zu schulischen Zwecken ist es jedoch zu komplex. Sollte die Lehrkraft den Schülern allerdings im Rahmen eines sprachwissenschaftlichen Seminars syntaktische Funktionen anhand von Baumdiagrammen näherbringen wollen, so kann das Korpus durchaus verwendet werden.

4.6.9.3. Semantisch annotierte Korpora: *SenSem* und *Internet with Semantic Roles*

Neben dem eben erwähnten Korpus AnCora (siehe Kapitel 2.5.2.3. und 4.6.9.2.) sind die Korpora *SenSem* sowie *Internet with Semantic Roles* semantisch annotiert (siehe Kapitel 2.5.3.). Das *SenSem* ist ein syntaktisch-semantisch annotiertes Korpus, dessen Besonderheit die Auflistung der Konkordanzen nach der jeweiligen Bedeutung des darin enthaltenen Verbs darstellt. Sofern der Nutzer nach einem Verb sucht, werden die verschiedenen Bedeutungen angezeigt und es können zu jeder Bedeutung Beispiele aufgerufen werden. Das *Internet with Semantic Roles* ist ein semantisch getaggttes Spezialkorpus, dessen Schwerpunkt im Vergleich zum Korpus *SenSem* auf der Suche nach semantischen Rollen liegt.

Die semantisch annotierten Korpora sind jedoch ebenso wie die syntaktisch annotierten für die Erreichung der Unterrichtsziele zu kompliziert, um von den Schülern genutzt zu werden, da diese im Spanischunterricht weder syntaktische noch semantische Sprachanalysen durchführen. Eine eingehende syntaktische oder semantische Untersuchung ist für einen kommunikativ ausgelegten Unterricht nicht produktiv, denn sie trägt trotz eines hohen Zeitaufwands nicht in ausreichendem Maß zur Erreichung des Ziels der Kommunikativität bei. Sollte die Semantik der spanischen Sprache jedoch in einem Kurs thematisiert werden, beispielsweise in einem speziellen wissenschaftspropädeutischen Seminar (W-Seminar) ¹²⁵⁰ oder einem anderen spezialisierten Spanischkurs, so sind die Korpora durchaus empfehlenswert, da sie frei zugänglich sind.

Alle in diesem Kapitel vorgestellten Korpora sind durch ihre Annotationen sehr wertvoll für die spanische Sprachwissenschaft, zu schulischen Zwecken sind jedoch vorwiegend die morphosyntaktisch getaggtten Korpora sinnvoll. Syntaktisch oder semantisch annotierte Korpora sind meist zu komplex, um sie in den Schulalltag einfließen zu lassen. Die Fachgebiete Syntax und Semantik sind in der Schule schließlich nicht als solche einzeln zu vertiefen, wie dies im Studium der Romanistik der Fall ist. Die syntaktisch und semantisch annotierten Korpora können interessierten Oberstufenschülern zwar für bestimmte Projekte vorgestellt werden, doch für den gewöhnlichen Spanischunterricht ist die Thematik zu schwierig. Im Gegensatz dazu kann

¹²⁵⁰ Vgl. Bayerisches Staatsministerium für Unterricht und Kultus (2007): op. cit. (01.05.2018).

die Lehrkraft die Korpora jedoch durchaus zur Ermittlung verschiedener Lesarten eines bestimmten Verbs oder zur Überprüfung der Existenz oder Korrektheit bestimmter syntaktischer Konstruktionen heranziehen.

4.6.10. Korpora auf CD-ROM: PRESEEA, LEXESP, C-ORAL-ROM und *Macrocorpus*

Die Verfügbarkeit der Korpora spielt eine wichtige Rolle, insbesondere im Hinblick auf die didaktische Nutzung derselben durch Lehrer und Schüler. Für den Unterricht sind diejenigen Korpora am besten geeignet, die im Internet kostenlos und unbegrenzt für jeden zugänglich sind, so wie dies beispielsweise bei den Korpora der *Real Academia Española* (siehe Kapitel 2.1.1.1., 2.1.1.2., 2.2.1.1. und 2.2.1.2.), aber auch bei vielen Spezialkorpora der Fall ist. Andere Korpora, wie diejenigen von Mark Davies (siehe Kapitel 2.1.1.3., 2.2.1.3. und 2.6.), sind zwar im Internet verfügbar, doch zum einen ist dafür eine Anmeldung erforderlich und zum anderen ist die Anzahl der Suchen darin bei einer Nutzung ohne Beitragszahlung begrenzt. Dennoch können alle diese Korpora im Unterricht genutzt werden, da sie im Internet zugänglich und zumindest in einem gewissen Rahmen kostenlos nutzbar sind. Sofern Korpora wie das PRESEEA (siehe Kapitel 2.8.2.1.), das LEXESP (2.8.2.2.), das *Macrocorpus* (2.4.2.1.) oder das C-ORAL-ROM (2.4.2.2.) allerdings nur auf CD-ROM erhältlich sind, kann nicht davon ausgegangen werden, dass sie für Lehrkräfte oder gar Schüler zugänglich sind. Aus diesem Grund wurden sie in der vorliegenden Arbeit nicht für die Erstellung von Unterrichtsmaterial berücksichtigt, da diese sich an alle Lehrkräfte richtet und daher die freie Verfügbarkeit der Korpora eine Grundvoraussetzung ist. Grundsätzlich können die auf CD-ROM verfügbaren Korpora ebenso gut verwendet werden, sodass einer Nutzung derselben, sofern erwünscht, nichts im Weg steht. Viele der beschriebenen Anwendungsmöglichkeiten können problemlos auf andere Korpora übertragen werden. Das C-ORAL-ROM als Korpus des gesprochenen Spanisch kann beispielsweise zur Erstellung von Aufgaben zum Hörverstehen, zur Schulung der Aussprache oder zur Erarbeitung kommunikativer Funktionen sowie anderer Besonderheiten der gesprochenen Sprache genutzt werden. Das Korpus LEXESP könnte als morphosyntaktisch annotiertes und lemmatisiertes Korpus der geschriebenen Sprache anstelle der Referenzkorpora für die Suche nach Wortarten, Lemmata oder Kookkurrenzen verwendet werden.

Ein großer Nachteil der Korpora auf CD-ROM ist neben der begrenzten Zugänglichkeit allerdings die Tatsache, dass diese auf dem Stand ihrer Veröffentlichung bleiben und nicht aktualisiert werden wie dies bei online zugänglichen Korpora meist der Fall ist. Daher sind letztere nicht nur aufgrund von deren Verfügbarkeit, sondern zudem wegen ihrer Aktualität zu schulischen Zwecken vorzuziehen.

Das Korpus PRESEEA ist zwar generell nur auf CD-ROM verfügbar, doch es sind einige Audiodateien und Transkripte im Internet kostenlos zugänglich. Im Fall des C-ORAL-ROM sind die ersten 30 Sekunden einiger Dateien im Internet anzuhören, welche sich aufgrund der zeitlichen Begrenzung jedoch nur bedingt für die Nutzung im Unterricht eignen. Auch das morphosyntaktisch annotierte LEXESP ist teilweise im Internet zugänglich, doch die Benutzeroberfläche ist nur spärlich ausgestattet, und die spanischen Sonderzeichen werden nicht korrekt dargestellt. Das *Macrocorpus* hingegen ist ausschließlich auf CD-ROM erhältlich.

Im Allgemeinen ist es für den schulischen Kontext wünschenswert, verschiedene Arten von Korpora zur Verfügung zu haben. Anhand allgemeiner Referenzkorpora können beispielsweise Vorkommen eines Ausdrucks ermittelt und untersucht werden, da sich hierin verschiedenste Texte aus unterschiedlichen Quellen zu vielschichtigen Themen finden. Ein Spezialkorpus, wie ein Zeitungskorpus, hingegen ermöglicht durch die Verfügbarkeit vollständiger Texte die Erstellung von Aufgaben zum Leseverstehen und anhand eines Korpus gesprochener Sprache können Aussagen zur Aussprache gemacht sowie spezielle Ausdrücke der gesprochenen Sprache erarbeitet werden. Ein pädagogisches Korpus wie das *Sacodeyl* bietet für den Einsatz im Unterricht besondere Vorteile, da die Texte hinsichtlich der Themen und des Niveaus kategorisiert sind, und multimodale Korpora bieten zusätzlich zu Text und Ton Videosequenzen, welche für die Schüler besonders interessant sind. Generell ist es jedoch sehr bedauerlich, dass der Großteil der Korpora keinen Zugang zu den Volltexten bietet.

Bei fortgeschrittenen Schülern in höheren Jahrgangsstufen kann durchaus mit unterschiedlichen Korpora gearbeitet werden, doch bei jüngeren Schülern empfiehlt es sich, nur wenige Korpora zu nutzen, deren Funktionsweise die Schüler nach ein paar Unterrichtsstunden beherrschen. Daher ist es durchaus von Vorteil, sich bei der Nutzung durch die Schüler auf die Verwendung der Korpora der *Real Academia Española* sowie

das CdE zu beschränken. Die Lehrkraft kann zur Erarbeitung von Unterrichtsmaterial, zur Vorbereitung auf den Unterricht oder für die Korrektur von Arbeiten allerdings sehr wohl verschiedene Korpora zurate ziehen.

4.7. Exkurs: Eigene Erstellung eines Korpus

Wie insbesondere in diesem Kapitel zu sehen war, kann die Arbeit mit den im Internet öffentlich zugänglichen Korpora sehr gewinnbringend für den Unterricht sein. Die Lehrkraft kann jedoch nicht nur bereits bestehende Korpora nutzen, sondern auch selbst ein Korpus zusammenstellen. Im Folgenden soll die Erstellung eines Korpus anhand eines Konkordanzprogramms beschrieben werden. Es handelt sich dabei um ein Spezialkorpus aus Texten, welche im Unterricht besprochen wurden, also eine Art pädagogisches Korpus. Ein sehr nützliches Programm zur Entwicklung eines eigenen Korpus ist *AntConc*¹²⁵¹, da es im Internet kostenlos heruntergeladen werden kann und zusätzlich zur Anzeige von Konkordanzen einige Funktionen bereithält, wie die Ermittlung von Kollokationen und *Clustern* (typischer Wortmuster), die Erarbeitung eigener Frequenz- und Wortlisten sowie die Arbeit mit getaggt Texten.

Es gäbe ebenfalls die Möglichkeit der Erstellung eines Korpus mit *SketchEngine*, einer Plattform für die Erstellung eigener und Nutzung von dort verfügbaren Korpora (siehe auch Kapitel 2.8.1.1.). Für eine Einzelmitgliedschaft ist jedoch eine monatliche Gebühr zwischen 4,83 € und 8,33 € zu entrichten, je nach der jeweiligen Nutzungsart.¹²⁵² Aufgrund der Kosten, die für die Erstellung eines Korpus mittels *SketchEngine* entstehen würden, soll diese Möglichkeit hier lediglich erwähnt werden. Da die vorliegende Arbeit jedoch öffentlich zugängliche Ressourcen nutzen möchte, welche von allen Privatpersonen kostenlos verwendet werden können, wird im Folgenden nicht weiter auf diese oder ähnliche Plattformen eingegangen, sondern die Erstellung eines Korpus anhand des kostenlos zur Verfügung stehenden Programms *AntConc* veranschaulicht werden.

¹²⁵¹ Vgl. Anthony (2018): *AntConc (Version 3.5.7) [Macintosh OS X]*. Op. cit. (14.05.2018).

¹²⁵² Vgl. Lexical Computing (ed.): *Sketch Engine. Price List*. Op. cit. (05.05.2018).

4.7.1. Einführung in das Konkordanzprogramm *AntConc*

Das eben erwähnte Programm *AntConc* ist von Laurence Anthony, einem Professor am *Center for English Language Education* an der Waseda Universität in Japan, entwickelt worden und wird konstant überarbeitet.¹²⁵³ Anthony stellt dieses sowie weitere von ihm entwickelte Programme auf seiner Homepage kostenfrei zur Verfügung, wobei verschiedene Versionen für *Windows*, *Macintosh OS X* oder *Linux* heruntergeladen werden können. Außerdem bietet er auf *Youtube* benutzerfreundliche Tutorials zu seinem Programm *AntConc*, welche für die Erstellung der in diesem Kapitel beschriebenen Korpora und der dazu folgenden Informationen und Hinweise genutzt wurden.¹²⁵⁴

Für die Erstellung eines Korpus müssen zunächst geeignete Dateien in das Programm geladen werden. Im vorliegenden Fall wird eine Textsammlung der Lehrwerke *¡Adelante! Nivel elemental*¹²⁵⁵ und *¡Adelante! Nivel intermedio*¹²⁵⁶ des Klett Verlags verwendet. Der Verlag bietet neben dem jeweiligen Schülerbuch ein Lehrerbuch an, das eine CD-ROM enthält, auf der alle Lehrbuchtexte digital enthalten sind.¹²⁵⁷ Das mit diesen Texten erstellte Korpus wird im Folgenden als *Adelante*-Korpus bezeichnet werden. Da das Programm *AntConc* mit *txt*-Dateien arbeitet, müssen die Texte des Lehrwerks *¡Adelante!* allerdings in *txt*-Dateien abgespeichert werden, bevor sie hochgeladen werden können.

Für gewöhnlich stellt das Umwandeln der *Word*-Dateien in *txt*-Dateien keine Probleme dar, manchmal kann es jedoch von Nutzen sein, Texte mit einem Textbearbeitungsprogramm zu modifizieren, beispielsweise wenn unnötige Ziffern,

¹²⁵³ Vgl. Anthony, Laurence (2016): *Laurence Anthony's Website – Contact Information*. Online Publikation: <http://www.laurenceanthony.net/contact.html> (07.06.2016).

¹²⁵⁴ Vgl. Anthony, Laurence (2010): *AntConc 3.2.4 – Tutorial 0: Getting Started*. Online Publikation: <https://www.youtube.com/watch?v=z9wwX7eR-Y> (15.02.2015). Es gibt für jede der beiden Programmversionen 3.2.4 und 3.4.0 insgesamt 11 beziehungsweise 10 Tutorials, welche jeweils verschiedene Funktionen des Programms genauer beschreiben. Die späteren Überarbeitungen, wie die hier verwendete, aktuelle Version 3.5.7 für *Macintosh*, basieren auf den Vorgängern, weshalb die Tutorials ebenso für die Arbeit mit den nachfolgenden Versionen verwendet werden können.

¹²⁵⁵ Vgl. Barquero, Antonio/ Collado Revestido, Cristina/ Corpas, Jaime/ Dienhoff, Oliver/ Domínguez, Yolanda/ Kuhlmann, Erika/ Navarro, Javier/ Reiter, Claudia (2010): *¡Adelante! Nivel elemental. Schülerbuch*. Stuttgart: Klett.

¹²⁵⁶ Vgl. Barquero/ Bizama/ Corpas/ Crovetto-Bizama/ Díaz Gutiérrez/ Jeske/ Jiménez Romera/ Navarro/ Pardellas Velay (2011): op. cit.

¹²⁵⁷ Vgl. Hueber, Margit/ Vega Ordoñez, Silvia (2010): *¡Adelante! Nivel elemental. Lehrerbuch mit Lehrersoftware und Video-DVD*. Klett: Stuttgart. Und: Huber, Margit/ Mancini, Lucia/ Simó Perales Zabka, Susana/ Vega Ordoñez, Silvia (2011): *¡Adelante! Nivel intermedio. Lehrerbuch mit Lehrersoftware und Video-DVD*. Stuttgart: Klett.

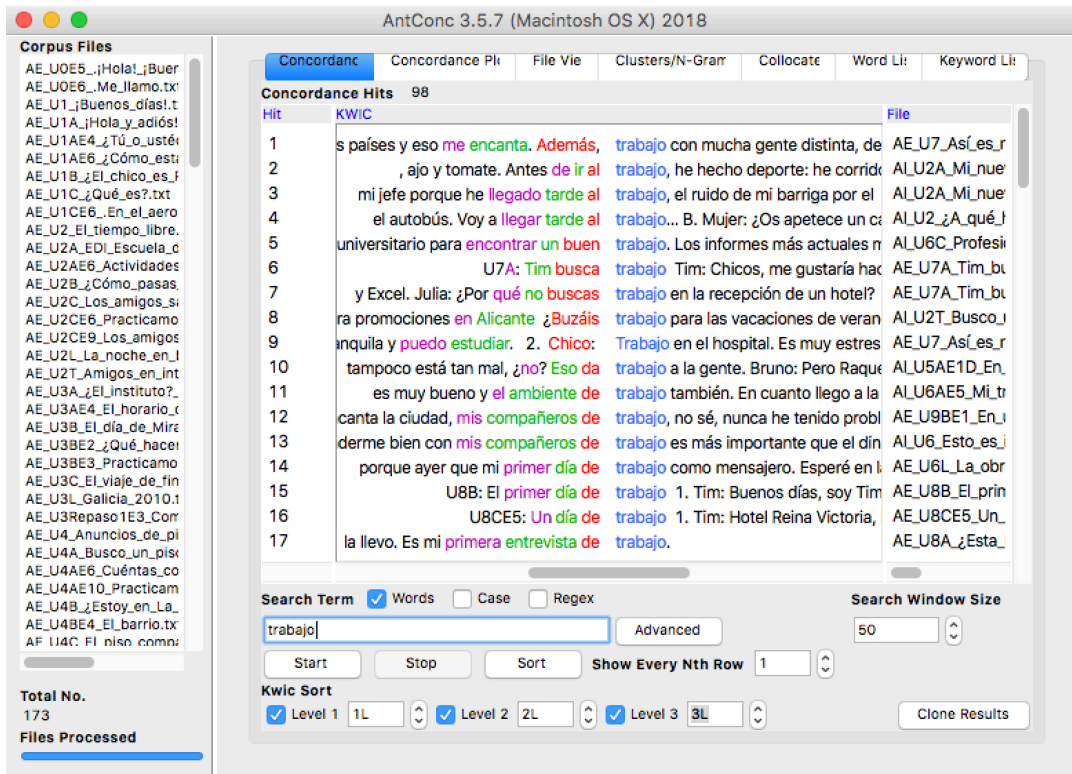
Markierungen, Aufzählungszeichen, Fußnoten oder Zwischenräume entfernt oder verändert werden sollen. Hierfür kann das Programm *jEdit* verwendet werden, sofern die gängigen Textbearbeitungsprogramme wie *Windows Word* nicht ausreichen. Mit *jEdit* können anhand von regulären Ausdrücken (siehe Kapitel 1.4.4.) verschiedenste Zeichenketten in Texten gesucht und durch andere Zeichen ersetzt oder gelöscht werden. Auch dieses Programm ist kostenlos im Internet verfügbar.¹²⁵⁸

4.7.2. Funktionen des Konkordanzprogramms *AntConc*

Das Programm *AntConc* verfügt über die Funktionen *Concordance*, *Concordance Plot*, *File View*, *Clusters/N-Grams*, *Collocates*, *Word List* und *Keyword List*, welche im Folgenden vorgestellt und näher erläutert werden. In der Funktion *Concordance* werden die Konkordanzen des Suchbegriffs angezeigt (siehe Kapitel 1.4.3.).¹²⁵⁹ Mittels der Option *Kwic Sort* können diejenigen Lexeme ausgewählt werden, welche farbig markiert werden sollen, beispielsweise die drei Lexeme links oder rechts des Suchbegriffs, wobei immer alle Lexeme in der Umgebung des Suchbegriffs angezeigt werden. Diese Option kann genutzt werden, um die gewöhnliche Verwendungsweise eines Lexems zu ermitteln. So liefert die Suche nach *trabajo* im *Adelante*-Korpus 98 Treffer, unter denen allein neun Mal der Ausdruck *encontrar (otro) trabajo* zu finden ist. Außerdem erscheinen die Kombinationen *ir al trabajo*, *llegar al trabajo*, *entrevista de trabajo*, *compañeros de trabajo* und *día de trabajo*.

¹²⁵⁸ Vgl. jEdit (ed.): *jEdit – Programmer's Text Editor*. Online Publikation: <http://jedit.org/index.php?page=download> (07.06.2016).

¹²⁵⁹ Für die Erstellung der folgenden Informationen zur Funktion *Concordance* wurde auf die Tutorials 1 und 2 der Version *AntConc 3.2.4* von Laurence Anthony zurückgegriffen: Vgl. Anthony, Laurence (2011): *AntConc 3.2.4 – Tutorial 1: Concordance Tool – Basic Features*. Online Publikation: <https://www.youtube.com/watch?v=9TsqFVrUYO0> (16.02.2015). Vgl. Anthony, Laurence (2011): *AntConc 3.2.4 – Tutorial 2: Concordance Tool – Advanced Features*. Online Publikation: <https://www.youtube.com/watch?v=pUN-flv-Cmw> (17.02.2015).

Abbildung 126: AntConc – Ergebnis der Suche nach *trabajo*.

Mithilfe der Option *Kwic Sort* können Schüler die bisher gelernten Verwendungsweisen des Vokabulars wiederholen und für die eigene Textproduktion nutzen. Dies kann vor allem dann von Vorteil sein, wenn Schüler für sie selbst schwierige und selten im Unterricht verwendete Lexeme in ihre eigenen Texte einbinden wollen und nicht wissen, wie diese in dieselben integriert oder mit welchen anderen Lexemen sie kombiniert werden können. Mittels der Arbeit mit Konkordanzen können Schüler auf den ersten Blick herausfinden, wie bestimmte Lexeme gebraucht werden, so zum Beispiel, dass man, anders als im Deutschen, das Adjektiv *otro* im Spanischen nicht mit dem unbestimmten Artikel kombiniert, wie an den Ausdrücken *encontrar otro trabajo* oder *en otro lugar* ersichtlich wird.

Im Programm *AntConc* sind jedoch nicht nur Suchen nach einzelnen Wortformen, sondern auch solche mit regulären Ausdrücken möglich (siehe Kapitel 1.4.4.). So kann nach *inca** gesucht werden, um alle Treffer für *inca* und *incas* zu sehen, oder nach *escrib**, um sich konjugierte Verbformen wie *escribo* und *escribían*, aber auch *escribir* und *escribiros* auflisten zu lassen, da der Asterisk auch in diesem Programm angibt, dass ein oder mehrere Zeichen nach dem gesuchten Lexem stehen können. Problematischer ist es bei unregelmäßigen Verbformen wie *dormir* – *duermo*. Hier muss separat nach *duerm**

gesucht werden. Ebenso können alle Gerundformen der Verben auf *-ar* mithilfe des Suchausdrucks *est* *ando* aufgelistet werden. Wenn man nun zusätzlich die Gerundformen für die Verben auf *-er* und *-ir* mittels des Suchausdrucks *est* *iendo* im Korpus ermittelt, kann aus den Konkordanzen ein Arbeitsblatt zur Wiederholung dieser Grammatikeinheit erstellt werden (vgl. Arbeitsblatt 25 *Repaso: Estar + gerundio*, Anhang 34). Diese Vorgehensweise ermöglicht es dem Lehrer, Arbeitsblätter zu bestimmten Themen zu erstellen, anhand derer die Schüler bereits gelernte Grammatikeinheiten mittels bekannter Sätze wiederholen oder die neue Grammatik induktiv erarbeiten können. Insbesondere leistungsschwächere Schüler sind sehr dankbar, wenn ihnen der Kontext von Grammatikübungen bereits bekannt ist, sodass sie sich auf die Einübung der Grammatik konzentrieren können. Durch den bekannten Kontext ist die Hemmschwelle niedriger und die Schüler verstehen nach derartigen Übungen die Grammatik in der Regel besser.

Außerdem bietet die Option *Advanced Search* die Möglichkeit, nach mehreren Lexemen gleichzeitig zu suchen, ebenso wie auch Listen von Lexemen aus einer abgespeicherten Datei in das Feld der *Advanced Search* geladen werden können. So kann beispielsweise gleichzeitig nach den in Kapitel 4.1.3.5. erwähnten Substantiven *posibilidad* und *oportunidad* gesucht werden. Als Ergebnis werden demnach alle Sätze mit einem der beiden Substantive aufgelistet, und die Schüler können mittels der Konkordanzen die Bedeutungsunterschiede anhand von ihnen bereits bekannten Texten ermitteln. Ebenso ist eine Suche nach *duro* im Kontext des Lexems *trabajo* möglich. Als Ergebnis werden somit nur diejenigen Sätze angezeigt, in denen sowohl *duro* als auch *trabajo* vorkommen, wobei die beiden Begriffe nicht unmittelbar benachbart sein müssen (wie es bei der gewöhnlichen Suche anhand der Eingabezeile der Fall ist), denn der Kontext kann in den Einstellungen der *Advanced Search* genau angegeben werden, beispielsweise als fünf Lexeme links und rechts des Suchbegriffs, was der Grundeinstellung entspricht.

Neben dem Reiter *Concordance* findet man den Reiter *Concordance Plot*, der dieselben Ergebnisse liefert, diese jedoch auf andere Weise darstellt.¹²⁶⁰ Hier wird jeder in das Programm geladene Text, der den Suchbegriff enthält, einzeln in einer Balkendarstellung aufgelistet und die darin enthaltene Trefferanzahl angegeben:

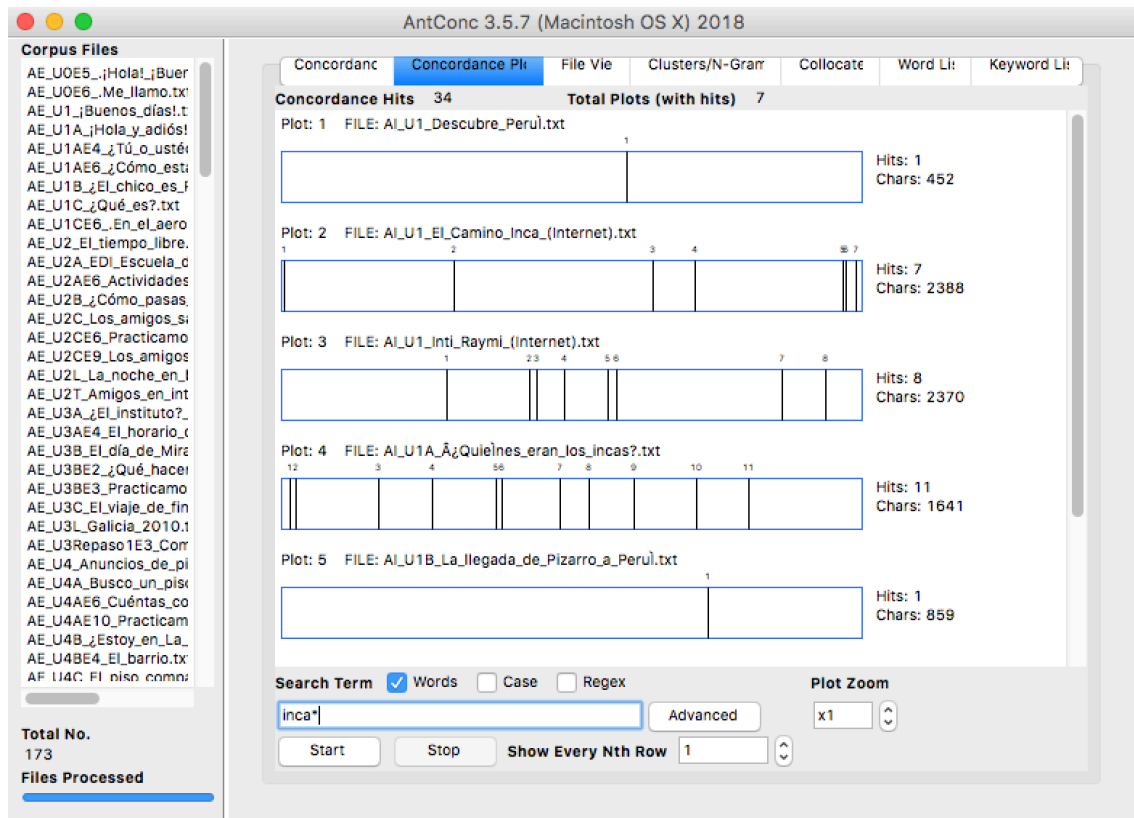


Abbildung 127: *AntConc* – Reiter *Concordance Plot* bei der Suche nach *inca**.

Die angezeigten Striche geben an, wo sich die Treffer im Text befinden. Somit kann sehr anschaulich dargestellt werden, welche Lexeme häufig in einer bestimmten Textsorte oder in Texten zu bestimmten Themen vorkommen und welche eher weniger oder gar nicht belegt sind. Jedoch ist bei der Interpretation der Ergebnisse Vorsicht geboten, da die Balken stets die gleiche Länge haben, ungeachtet dessen, wie viele Lexeme der Text umfasst. Die Anzahl der Lexeme im Text ist allerdings zusammen mit der Summe der Treffer für den jeweiligen Suchbegriff auf der rechten Seite angegeben. Alle Möglichkeiten zur Formulierung von Suchausdrücken im Reiter *Concordance* sind

¹²⁶⁰ Für die Erstellung der folgenden Informationen zur Funktion *Concordance Plot* wurde auf das Tutorial 3 der Version *AntConc 3.2.4* von Laurence Anthony zurückgegriffen: Vgl. Anthony, Laurence (2011): *AntConc 3.2.4 – Tutorial 3: Concordance Plot Tool – Basic Features*. Online Publikation: <https://www.youtube.com/watch?v=YNDvcxonmP4> (19.02.2015).

hier ebenso gegeben, so kann auch mit regulären Ausdrücken oder mittels der Option *Advanced Search* gearbeitet werden. Mithilfe der Funktion *Plot Zoom* können die eben erwähnten Balkendarstellungen für die jeweiligen Texte näher herangezoomt werden, wodurch sich genauer feststellen lässt, wo sich die Treffer im Text befinden. Durch einen Klick auf eine der Linien gelangt der Benutzer in die Ansicht *File View* und sieht den gesamten Text, in dem der Suchausdruck vorkommt.

Die Funktion *Concordance Plot* ist bei einem Korpus mit Lehrwerkstexten zwar recht interessant, erscheint jedoch vor allem im Hinblick auf ein Lernerkorpus relevant (siehe Kapitel 2.2.2.), da der Lehrer oder auch die Schüler so auf einen Blick sehen, ob und wann sich bestimmte Lexeme in den eigenen Texten häufen und wie die Schüler ihre schriftlichen Textproduktionen verbessern können.

Im Reiter *File View* werden die ins Programm geladenen Texte, wie bereits erwähnt, vollständig angezeigt.¹²⁶¹ Es können in dieser Funktion einzelne Dateien ausgewählt und darin nach bestimmten Lexemen gesucht werden. Auf diese Weise sieht man innerhalb kürzester Zeit, wie oft beispielsweise das Substantiv *hambre* im ausgewählten Dokument erscheint:

¹²⁶¹ Für die Erstellung der folgenden Informationen zur Funktion *File View* wurde auf das Tutorial 4 der Version *AntConc* 3.2.4 von Laurence Anthony zurückgegriffen: Vgl. Anthony, Laurence (2011): *AntConc* 3.2.4 – *Tutorial 4: File View Tool – Basic Features*. Online Publikation: https://www.youtube.com/watch?v=Yalioyw_74M (30.03.2015).

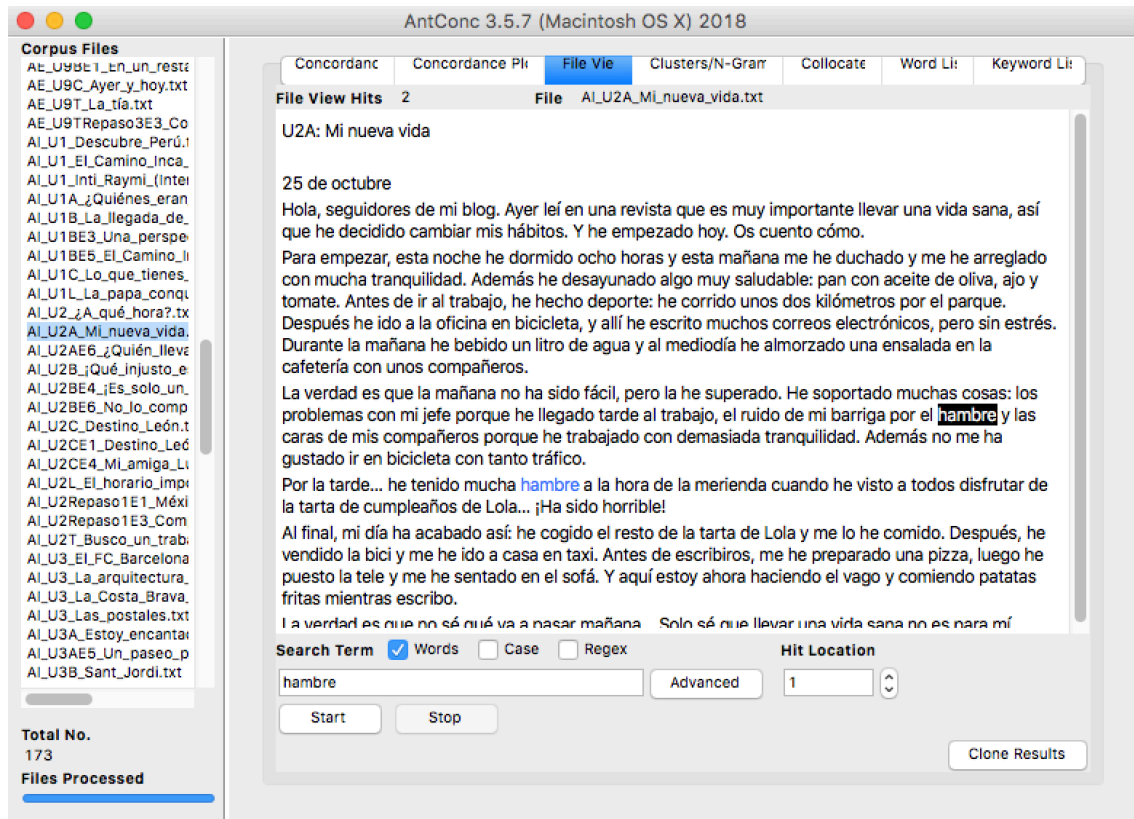


Abbildung 128: AntConc – File View bei der Suche nach *hambre* in einem ausgewählten Text.

Mittels der Option *Hit Location* am unteren Rand des Fensters kann zusätzlich im Text von einem Treffer zum nächsten gesprungen werden. Auch hier sind die Optionen der *Advanced Search* möglich, sodass herausgefunden werden kann, welche Lexeme zusammen mit anderen Schlagwörtern im Text vorkommen, wobei diese gleich im Textzusammenhang angezeigt werden. Wenn man auf einen blau geschriebenen Treffer klickt, wechselt das Programm in den Reiter *Concordance* und alle Treffer werden untereinander aufgelistet. Wie zu sehen ist, sind alle Funktionen miteinander verlinkt, wobei die Funktion *File View* im Zentrum steht.

Die Funktion *Clusters* ermöglicht es dem Benutzer, für Suchbegriffe typische Muster zu ermitteln.¹²⁶² Es können beispielsweise *encontrar trabajo* oder *trabajo en* bei

¹²⁶² Für die Erstellung der folgenden Informationen zur Funktion *Clusters/N-Grams* wurde auf die Tutorials 5 und 6 der Version *AntConc 3.2.4* sowie auf die Tutorials 6 und 7 der Version *AntConc 3.4.0* von Laurence Anthony zurückgegriffen:

Vgl. Anthony, Laurence (2011): *AntConc 3.2.4 – Tutorial 5: Clusters Tool – Basic Features*. Online Publikation: <https://www.youtube.com/watch?v=5kosY5xIsLE> (01.04.2015).

Vgl. Anthony, Laurence (2011): *AntConc 3.2.4 – Tutorial 6: N-Grams Tool – Basic Features*. Online Publikation: <https://www.youtube.com/watch?v=-vv663HjTso> (02.04.2015).

der Suche nach *trabajo* gefunden werden, wobei die Größe der *Cluster* variiert werden kann und die Treffer nach abfallender Häufigkeit angeordnet werden. Auch hier kann wieder angegeben werden, in welchem Intervall links und rechts des Lexems gesucht werden soll. Zudem kann bestimmt werden, ob der Suchbegriff innerhalb des *Clusters* rechts oder links stehen soll. Neben der Auflistung der Ergebnisse nach Häufigkeit oder Alphabet können die Treffer auch nach der Wahrscheinlichkeit, mit welcher der Suchbegriff zusammen mit dem direkt darauffolgenden Lexem erscheint, sortiert werden. Die Wahrscheinlichkeit 1.00 bedeutet, dass dieses Lexem im Korpus immer, also zu 100% mit dem anderen erscheint, und 0.50, dass die Bestandteile des *Clusters* in jedem zweiten Fall zusammen vorkommen. Bei der Suche nach *piso* ermittelt *AntConc* für das *Adelante*-Korpus beispielsweise, dass es in 20,8% der Fälle von *compartido* gefolgt wird. Zudem finden sich in der Trefferliste weitere interessante Kollokationen, wie *piso barato*, *piso céntrico* oder *piso tranquilo*, welche jedoch jeweils nur einmal vorkommen. Diese und ähnliche Suchen sind vor allem in umfassenderen Korpora von Bedeutung, sie können aber auch in kleineren Korpora für schulische Zwecke aufschlussreiche Ergebnisse liefern.

Außerdem besteht in der Funktion *Clusters* die Möglichkeit, nach sogenannten N-Grammen zu suchen (siehe auch Kapitel 1.3.2.1.). Sofern man diese Option auswählt, kann kein Suchbegriff mehr in die Leiste eingegeben werden, denn das Programm sucht nach Kombinationen mit der angegebenen Anzahl von Lexemen. Die Grundeinstellung ist zwei Lexeme, aber vor allem bei einer größeren Lexemanzahl wird der Sinn dieser Option klar:

Vgl. Anthony, Laurence (2014): *AntConc 3.4.0 – Tutorial 6: Clusters Tool*. Online Publikation: <https://www.youtube.com/watch?v=DV8YC0zKBOg> (08.09.2016).

Vgl. Anthony, Laurence (2014): *AntConc 3.4.0 – Tutorial 7: N-Grams Tool*. Online Publikation: <https://www.youtube.com/watch?v=1FG9N3jqlHY> (08.09.2016).

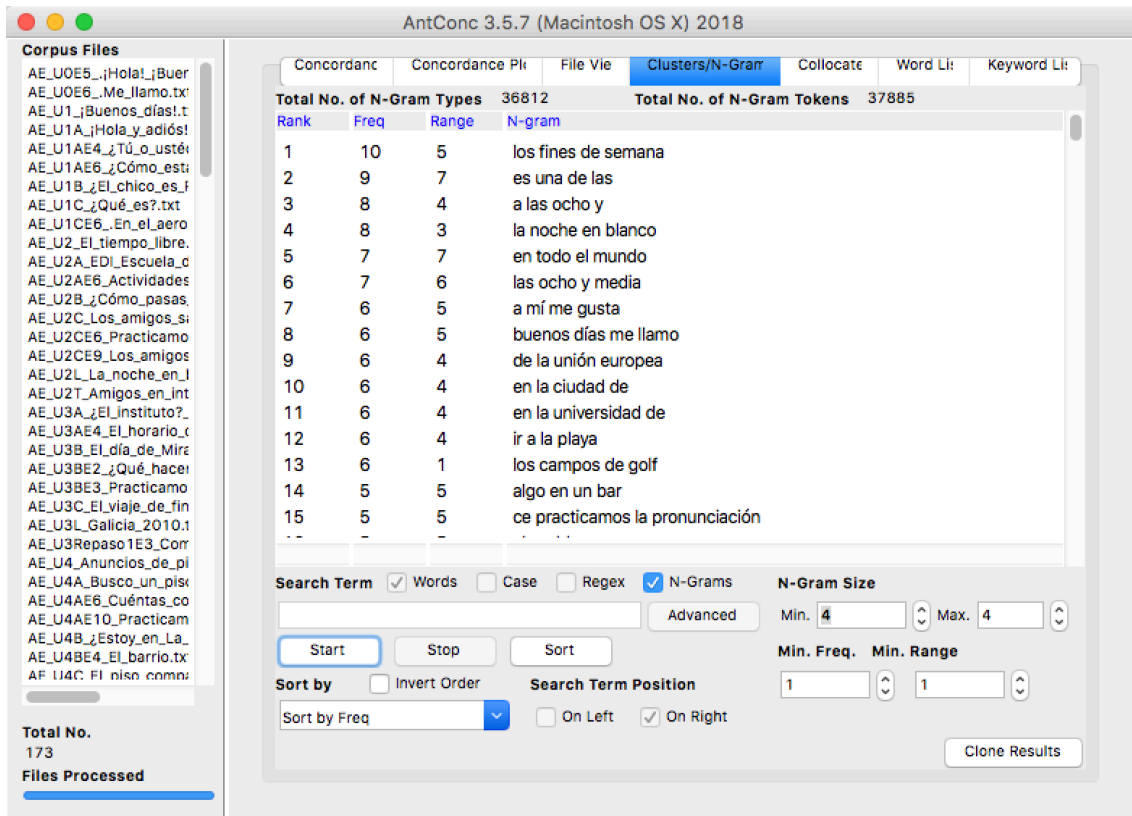


Abbildung 129: AntConc – Funktion Clusters bei der Suche nach N-Grams.

Als Ergebnisse für N-Gramme mit vier Lexemen finden sich im *Adelante*-Korpus unter anderem gebräuchliche Ausdrücke wie *los fines de semana*, *en todo el mundo* oder *los campos de golf*, die je nach Textauswahl auch schon in kleinen Textsammlungen deutlich zu erkennen sind. Diese Ausdrücke können für Lerner einer Fremdsprache sehr hilfreich sein, vor allem, wenn diese vom Lehrwerk nicht als eigenständige Ausdrücke, sondern nur als Einzelbestandteile vermittelt werden.

Eine weitere sehr nützliche Funktion findet sich im Reiter *Collocates*, das dem Benutzer anzeigt, welche Lexeme eng mit dem Suchbegriff verbunden sind.¹²⁶³ Es kann mittels dieser Funktion nach allen im Korpus vorkommenden Kollokationen zum eingegebenen Lexem gesucht werden, wobei in der Trefferliste neben dem jeweiligen Kollokator auch die Rangnummer, die gesamte Häufigkeit und das Vorkommen links und rechts des Suchbegriffs aufgelistet werden:

¹²⁶³ Für die Erstellung der folgenden Informationen zur Funktion *Collocates* wurde auf das Tutorial 7 der Version *AntConc 3.2.4* von Laurence Anthony zurückgegriffen: Vgl. Anthony, Laurence (2011): *AntConc 3.2.4 – Tutorial 7: File Collocates Tool – Basic Features*. Online Publikation: <https://www.youtube.com/watch?v=An6gb1W3pFM> (02.04.2015).

4. Nutzung von Korpora in der fremdsprachlichen Unterrichtspraxis

AntConc 3.5.7 (Macintosh OS X) 2018

Corpus Files

- AE_U0E5_¡Hola!.txt
- AE_U0E6_Me llamo.txt
- AE_U1_¡Buenos días!.txt
- AE_U1A_¡Hola y adiós!
- AE_U1AE4_¿Tú o usted?
- AE_U1AE6_¿Cómo está?
- AE_U1B_¿El chico es?
- AE_U1C_¿Qué es? .txt
- AE_U1CE6_En el aeropuerto
- AE_U2_¿El tiempo libre?
- AE_U2A_¿Escuela?
- AE_U2AE6_Actividades
- AE_U2B_¿Cómo pasas?
- AE_U2C_¿Los amigos?
- AE_U2CE6_Practicamos
- AE_U2CE9_¿Los amigos?
- AE_U2L_¿La noche en?
- AE_U2T_¿Amigos en?
- AE_U3A_¿El instituto?
- AE_U3AE4_¿El horario?
- AE_U3B_¿El día de?
- AE_U3BE2_¿Qué hacer?
- AE_U3BE3_Practicamos
- AE_U3C_¿El viaje de?
- AE_U3L_Galicia 2010.
- AE_U3Repasso1E3_Corr
- AE_U4_Auncios de?
- AE_U4A_Busco un?
- AE_U4AE6_Cuéntas?
- AE_U4AE10_Practicamos
- AE_U4B_¿Estoy en?
- AE_U4BE4_¿El barrio?
- AE_U4C_¿El niño com?

Total No. 173
Files Processed

Concordance Concordance Pl File Vie Clusters/N-Gram **Collocates** Word List Keyword List

Total No. of Collocate Types: 67 Total No. of Collocate Tokens: 135

Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate
1	1	0	1	10.83665	une
2	1	1	0	10.83665	mítico
3	1	0	1	10.83665	garcilaso
4	1	0	1	10.83665	ficticio
5	1	0	1	10.83665	escribían
6	1	1	0	10.83665	cronista
7	1	0	1	10.83665	construían
8	1	0	1	10.83665	conocían
9	1	0	1	10.83665	comían
10	1	0	1	10.83665	andas
11	13	11	2	10.14477	camino
12	1	1	0	9.83665	hablaban
13	2	1	1	9.83665	emperador
14	1	0	1	9.83665	cultivaban
15	2	1	1	9.25169	raymi
16	1	0	1	9.25169	online
17	1	0	1	9.25169	nuevos

Search Term ☒ Words ☐ Case ☐ Regex

inca* Advanced From... 2L To... 2R

Start Stop Sort

Sort by ☐ Invert Order

Sort by Stat

Window Span ☐ Same

Min. Collocate Frequency 1

Clone Results

Abbildung 130: AntConc – Funktion Collocates bei der Suche nach *inca**.

Anhand der angezeigten statistischen Angaben kann beispielsweise ermittelt werden, dass die Lexeme *mítico*, *garcilaso*, *camino* oder auch *emperador* stärker mit dem Suchbegriff *inca* verbunden sind als *de*, *a* oder *un* (ganz am Ende der Liste), da die Werte der erstgenannten über 9.0 und die Werte für letztere lediglich bei ungefähr 2.0 liegen. Unter den signifikantesten Kollokatoren werden allerdings auch einige Verbformen, meist im *pretérito imperfecto*, aufgelistet, da diese Zeitform im Lehrwerk *¡Adelante! Nivel intermedio* in der Lektion über Peru eingeführt beziehungsweise wiederholt wird. Da die Verbformen ausschließlich in Zusammenhang mit den Inka im Lehrbuch vorkommen, werden sie hier als besonders signifikant aufgelistet. Die Funktion *Collocates* kann jedoch besser unter Verwendung einer Lemmaliste genutzt werden, da in diesem Fall alle *tokens*, also die einzelnen Wortformen, eindeutig der jeweiligen Grundform zugeordnet werden (siehe Kapitel 1.4.1.). Im nächsten Abschnitt über die Funktion *Word List* wird genauer auf die Arbeit mit einer Lemmaliste eingegangen.

Außerdem wird neben der allgemeinen Frequenz im Korpus auch angezeigt, wie häufig das jeweilige Lexem links und rechts des Suchbegriffs erscheint. Diese Auskunft über die Wortstellung kann für Lehrer und Schüler ebenfalls sehr aufschlussreich sein,

insbesondere bei Adjektiven im Hinblick auf deren Position vor oder nach Substantiven. Die Ergebnisse werden in der Grundeinstellung nach der Stärke der Kollokation sortiert, wobei die stärkste Kollokation ganz oben angezeigt wird. Der hierbei herangezogene Maßstab zur Ermittlung der Kollokationen ist der Wert der MI, es kann jedoch ebenso *T-Score* als Maßstab ausgewählt werden (siehe Kapitel 1.4.8.).

Im Reiter *Word List* sind alle im Korpus enthaltenen Lexeme mit ihrer Rangnummer und ihrer Häufigkeit im Korpus aufgelistet.¹²⁶⁴ Wenn man in diesem Reiter auf *Start* klickt, wird zunächst eine Liste der häufigsten Lexeme erstellt. In der oberen Leiste ist die Gesamtzahl der *types* und *tokens* (siehe Kapitel 1.4.1.) ersichtlich und damit auch die Größe des Korpus. Im *Adelante*-Korpus handelt es sich beispielsweise um 5527 *types* und 38404 *tokens*. Die Ergebnisse können nach Häufigkeit, alphabetisch nach dem Anfang oder auch nach dem Ende des Lexems sortiert werden. Zudem gibt es in den *Tool Preferences* der *Word List* die Möglichkeit, frequente Funktionswörter wie beispielsweise *de*, *con*, *un*, *la* oder *el* mittels einer *Stoplist* auszuschließen. Des Weiteren kann hier eine Lemmaliste hochgeladen werden, wobei die Arbeit mit derselben es dem Benutzer ermöglicht, nicht nur die Häufigkeit eines Lexems zu sehen, sondern auch die damit verwandten Formen. Beispielsweise kann für das Verb *ser* nicht nur die Häufigkeit des Infinitivs, sondern die gesamte Häufigkeit aller verschiedenen Formen von *ser*, also die Anzahl aller Wortformen wie *soy*, *eres* oder *era* angezeigt werden, wie im folgenden Anwendungsbeispiel zu sehen sein wird.

In einem Kurs der 12. Jahrgangsstufe für Spanisch als spät beginnende Fremdsprache werden verschiedene Texte aus unterschiedlichen Quellen zur Bearbeitung des Themengebiets Argentinien herangezogen.¹²⁶⁵ Um nun entweder den Schülern eine

¹²⁶⁴ Für die Erstellung der folgenden Informationen zur Funktion *Word List* wurde auf das Tutorial 8 der Version *AntConc 3.2.4* von Laurence Anthony zurückgegriffen: Vgl. Anthony, Laurence (2011): *AntConc 3.2.4 – Tutorial 8: Word List Tool – Basic Features*. Online Publikation: https://www.youtube.com/watch?v=Zb71yaBP_II (07.04.2015).

¹²⁶⁵ Bei den Texten, die für die Unterrichtseinheit zum Thema Argentinien verwendet werden, handelt es sich um folgende Quellen:

Amann, Klaus/ Marín Barrera, Sara/ Schleyer, Jochen/ Vicente Álvarez, Araceli/ Wlasak-Feik, Christine (2007): *Encuentros 2. Método de español*. Berlin: Cornelsen, S. 54.

Escárate López, Luis M./ Steveker, Wolfgang/ Vences, Ursula/ Wlasak-Feik, Christine (2014): *Punto de vista. Nueva Edición. Texte und Aufgaben für die Oberstufe*. Berlin: Cornelsen, S. 164, 184-193.

Petry, Kristina (2015): *RAAbits Spanisch. Grundwerk*. Stuttgart: Dr. Josef Raabe Verlag, M6/M12, S. 14, 26.

4. Nutzung von Korpora in der fremdsprachlichen Unterrichtspraxis

Liste der frequentesten Lexeme zur Vorbereitung auf das Thema an die Hand zu geben oder das noch unbekannte Vokabular für eine Klausur zu bestimmen, kann die Arbeit mit einem speziell aus diesen Texten zusammengestellten Korpus förderlich sein. Anhand des Korpus aus diesen Texten – im vorliegenden Fall 14 Texte rund um die Kultur Argentiniens – können die Lehrkraft oder auch Schüler bereits einiges über häufig verwendete Lexeme herausfinden, sich über bestimmte Schlagworte informieren oder nach interessanten sprachlichen Strukturen suchen. Noch aussagekräftiger ist in Bezug auf die Häufigkeit von Lexemen allerdings die Arbeit mit einer Lemmaliste. Eine derartige Lemmaliste für die spanische Sprache steht auf Laurence Anthonys Homepage unter der Bezeichnung *Spanish Lemma List* kostenlos zum Download zur Verfügung.¹²⁶⁶

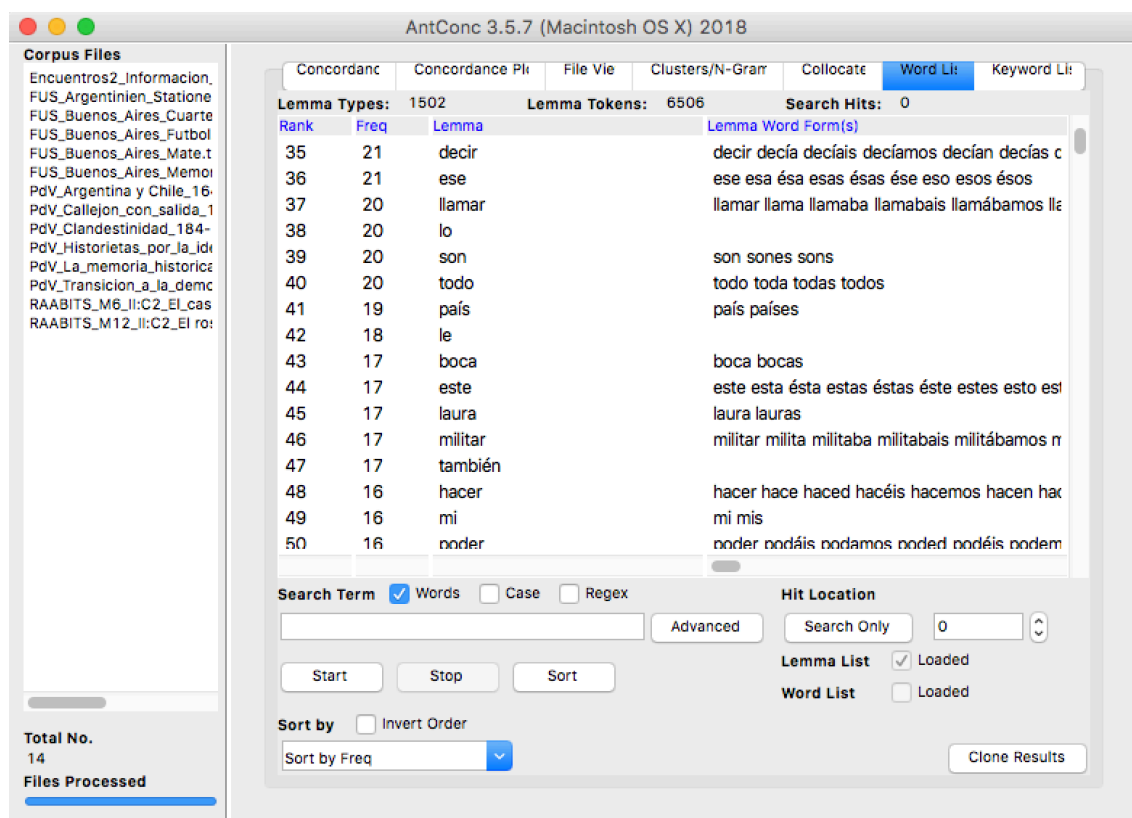


Abbildung 131: *AntConc* – *Word List* des *Argentina*-Korpus unter Verwendung einer Lemmaliste.

Ohne die Lemmaliste wird bei *AntConc* lediglich die Anzahl der Vorkommen der jeweiligen Wortform, also die Zahl der *tokens* für diesen einen *type*, angegeben, ohne

Schütz, Susanne (2013): „Un paquete de Buenos Aires. Mehrkanaliges Lernen an landeskundlichen Stationen“. In: *Der Fremdsprachliche Unterricht Spanisch – Metrópolis*, 2013, Band 40, S. 30-35, inklusive CD-ROM.

¹²⁶⁶ Vgl. Anthony, Laurence (2018): *Spanish Lemma List*. Online Publikation: <http://www.laurenceanthony.net/software/antconc/> (17.05.2018).

jedoch zusammengehörige Formen eines Verbs oder Singular- und Pluralformen eines Substantivs zusammenzufassen. Bei der Arbeit mit der Lemmaliste werden die Wortformen jedoch einem Lemma zugeordnet und zusammengezählt, wie in Abbildung 131 zu sehen ist.

Der letzte verfügbare Reiter im Eingabefenster von *AntConc* ist die *Keyword List*.¹²⁶⁷ Mittels dieser Funktion ist es dem Benutzer möglich, einen Text oder ein Korpus mit einem Referenzkorpus zu vergleichen, um herauszufinden, welche Lexeme in Ersterem ungewöhnlich häufig vorkommen. Als Methoden für die Auswahl der Schlüsselwörter können sowohl der *Log-Likelihood-Test* als auch der *Chi-Quadrat-Test* ausgewählt werden. Beide Methoden können zur Ermittlung der Signifikanz herangezogen werden, sie prüfen also, ob mit genügend großer Wahrscheinlichkeit davon ausgegangen werden kann, dass diese Lexeme im untersuchten Korpus häufiger vorkommen als im Referenzkorpus.¹²⁶⁸ Hierbei muss jedoch beachtet werden, dass der *Log-Likelihood-Test* gegenüber niedrigen Werten robuster ist. Im *Chi-Quadrat-Test* hingegen wird bei niedrigen Werten schnell Signifikanz angezeigt, obwohl anhand der Datenmenge nicht klar ist, ob diese wirklich vorliegt.¹²⁶⁹ Der Ersteller von *AntConc*, Laurence Anthony, empfiehlt die Verwendung der *Log-Likelihood-Methode*, weshalb diese auch der Grundeinstellung entspricht.

Sofern also in den *Tool Preferences* ein Vergleichskorpus hochgeladen wird, kann unter Zuhilfenahme einer zuvor automatisch generierten *Word List* eine *Keyword List* zusammengestellt werden, welche signifikant häufiger verwendete Lexeme des Vergleichstexts zusammen mit einer Art Häufigkeitsindex (*Keyness*) angibt. Außerdem ist es möglich, sich negative *Keywords*, also Lexeme, die auffällig selten verwendet werden, anzeigen zu lassen. Hierfür ist lediglich die Option *Show negative keywords* auszuwählen.

¹²⁶⁷ Für die Erstellung der folgenden Informationen zur Funktion *Keyword List* wurde auf das Tutorial 9 der Version *AntConc* 3.2.4 von Laurence Anthony zurückgegriffen: Vgl. Anthony, Laurence (2011): *AntConc 3.2.4 Tutorial 9: Keyword List Tool - Basic Features*. Online Publikation: <https://www.youtube.com/watch?v=JvasTvQY7kU> (22.07.2015).

¹²⁶⁸ Vgl. Bubenhofer, Noah: *Statistik: Einführung*. Online Publikation: http://www.bubenhofer.com/korpuslinguistik/kurs/index.php?id=statistik_einfuehrung.html (06.09.2016).

¹²⁶⁹ Vgl. Bubenhofer, Noah: *Induktive Statistik: Signifikanztest (2)*. Online Publikation: http://www.bubenhofer.com/korpuslinguistik/kurs/index.php?id=statistik_signifikanzLLR.html (06.09.2016).

Für die Erstellung einer *Keyword List* kann nicht nur ein Vergleichskorpus in Form von Texten verwendet werden, sondern auch die *Word List* eines solchen Korpus. In dieser *Word List* sollten die vorkommenden Lexeme nach Rang und unter Angabe von deren gesamter Häufigkeit aufgelistet sein. Die *Real Academia Española* bietet zusätzlich zu deren Korpus CREA (Version 2015, siehe Kapitel 2.1.1.1.) auch Listen der häufigsten 1000, 5000, 10000 oder sogar eine Liste aller im Korpus vorkommenden Lexeme an.¹²⁷⁰ Diese Listen können als Frequenzliste im Programm *AntConc* fungieren.

Als Grundlage für die folgenden Anwendungsmöglichkeiten wird die eben erwähnte Liste aller im Korpus CREA (Version 2015) verwendeten Lexeme mit deren Häufigkeit ausgewählt. Da diese Liste jedoch formal nicht den Anforderungen von *AntConc* entspricht und eine Fehlermeldung angezeigt würde, sofern man sie in das Programm lädt, sollte man sie zunächst mithilfe der Programme *jEdit* und *Excel* bearbeiten, um sie für das Programm kompatibel zu gestalten. Eine für *AntConc* kompatible Liste sollte den Rang als Kardinal- und nicht als Ordinalzahl, das jeweilige Lexem sowie dessen absolute Häufigkeit enthalten, jedoch über keine Angabe zur normalisierten Häufigkeit verfügen, wie dies in der ursprünglichen Liste der Fall ist.

Wenn man nun eine *Keyword List* erstellt, kann man erkennen, dass neben Funktionswörtern beispielsweise die Lexeme *mate*, *argentina* und *boca* in den Texten des *Argentina*-Korpus überproportional häufig repräsentiert sind. Die Auflistung der Funktionswörter erschwert allerdings das Erkennen der für diese spezielle Thematik häufigen und dadurch wichtigen Lexeme. Um die Ergebnisse der Häufigkeitslisten aussagekräftiger zu gestalten, ist die Arbeit mit einer sogenannten *Stoplist* ratsam. In dieser Liste sind Lexeme enthalten, die allgemein sehr häufig vorkommen. Es handelt sich vorwiegend um Funktionswörter, wobei auch sehr frequente und vielseitig gebrauchte Verben und Substantive darin vorkommen können, so beispielsweise *deber*, *decir*, *estar*, *cosa* oder *lugar*. Die *Stoplist* kann entweder selbst erstellt oder im Internet heruntergeladen werden. Für die Verwendung im *Argentina*-Korpus wird eine Stopliste aus dem Internet genutzt, die jedoch um ein paar Personalpronomen und Possessivbegleiter erweitert wurde, welche ohne ersichtlichen Grund fehlten.¹²⁷¹ Die

¹²⁷⁰ Vgl. Real Academia Española (ed.): *CREA – Listado de frecuencias*. Online Publikation: <http://corpus.rae.es/lfrecuencias.html> (22.07.2015).

¹²⁷¹ Vgl. Bougé, Kevin: *Download stop words*. Online Publikation: <https://sites.google.com/site/kevinbouge/stopwords-lists> (12.02.2016).

mittels der Stopliste ermittelte *Keyword List* lässt nun auf den ersten Blick erkennen, dass unter anderem die Lexeme *mate*, *argentina*, *años*, *boca*, *américa* und *mayo* Schlüsselwörter dieses Korpus sind:

AntConc 3.5.7 (Macintosh OS X) 2018

Corpus Files

- Encuentros2_Informacion
- FUS_Argentinien_Stations
- FUS_Buenos_Aires_Cuarte
- FUS_Buenos_Aires_Futbol
- FUS_Buenos_Aires_Mate.t
- FUS_Buenos_Aires_Memori
- PdV_Argentina y Chile_16
- PdV_Callejon_con_salida_1
- PdV_Clandestinidad_184-
- PdV_Historietas_por_la_idi
- PdV_La_memoria_historica
- PdV_Transicion_a_la_democ
- RAABITS_M6_II:C2_El_cas
- RAABITS_M12_II:C2_El ro

Keyword Types: 719 Keyword Tokens: 1707 Search Hits: 0

Rank	Freq	Keyness	Effect	Keyword
1	23	+ 496.91	0.0147	mate
2	59	+ 473.19	0.0052	argentina
3	19	+ 410.47	0.0121	años
4	17	+ 367.25	0.0109	boca
5	16	+ 345.64	0.0102	américa
6	14	+ 302.43	0.0089	mayo
7	14	+ 302.43	0.0089	mundo
8	13	+ 280.82	0.0083	democracia
9	11	+ 237.61	0.007	casa
10	11	+ 237.61	0.007	ramón
11	10	+ 216.01	0.0064	agua
12	10	+ 216.01	0.0064	pepón
13	30	+ 211.55	0.0031	fútbol
14	9	+ 194.4	0.0058	abuelas
15	8	+ 172.8	0.0051	madre
16	8	+ 172.8	0.0051	mundial
17	22	+ 155.28	0.0029	aires

Search Term: ☒ Words ☐ Case ☐ Regex Hit Location: Advanced Search Only: 0

Start Stop Sort Reference Corpus: ☒ Loaded Clone Results

Sort by: ☐ Invert Order Sort by Keyness

Total No. 14 Files Processed

Abbildung 132: *Argentina*-Korpus – Ermittlung einer *Keyword List* unter Verwendung einer Stopliste.

Die Ermittlung einer *Keyword List* bietet dem Lehrer die Möglichkeit, den Schülern zu jedem Themenbereich eine Vokabelliste mit tatsächlich benötigtem Vokabular an die Hand zu geben. Diese Liste kann auf verschiedene Weise eingesetzt werden:

- zur Vorbereitung des Vokabulars für das folgende Thema,
- zum Einstieg als Hilfestellung für eine Brainstormingphase,
- während der Bearbeitung des Themengebiets zur Anfertigung einer Mindmap,
- als Abschluss zur Wiederholung des Inhalts anhand des darin vorkommenden Vokabulars.

Im Oberstufenkurs Spanisch als spätbeginnende Fremdsprache wurde den Schülern im dritten Lernjahr vor der Bearbeitung des Themengebiets Argentinien eine kurze Vokabelliste an die Hand gegeben (siehe Arbeitsblatt 26 *Argentina – Vocabulario más frecuente*, Anhang 35). Die Schüler erhielten den Auftrag, die Bedeutung dieser Lexeme bis zur nächsten Stunde zu ermitteln, sofern sie ihnen noch unbekannt waren. Auf diese

Weise wird zusätzlich zur Vokabelvorentlastung auch eine Individualisierung und Differenzierung erreicht, da die Schüler sich lediglich mit den für sie unbekannten Lexemen befassen müssen, um den Großteil der Texte zu verstehen.

In der Aufzählung, deren Anfang in Abbildung 132 zu sehen ist, finden sich auch Namen von Charakteren aus bestimmten Texten (beispielsweise Ramón und Laura), doch diese Tatsache verdeutlicht, dass es sich um eine Liste handelt, welche nach der Frequenz des Vokabulars sortiert, aus dem Programm entnommen und nicht vom Lehrer verändert wurde. Sofern die *Keyword List* am Ende des Themengebiets als Abrundung verwendet wird, können die Eigennamen dazu beitragen, sich die einzelnen Texte wieder ins Gedächtnis zu rufen und daraufhin eine kurze Zusammenfassung zu geben.

Auf ähnliche Weise wurde in der darauffolgenden Lektion zum Thema Chile verfahren.¹²⁷² Hier wurde das Arbeitsblatt mit der *Keyword List* zum Themenabschluss für die Methode ‚Blitzlicht‘ verwendet. Die Schüler erhielten das Arbeitsblatt 27 *Chile – Vocabulario más frecuente* (Anhang 36) und hatten drei Minuten Zeit, um sich die Lexeme durchzulesen und sich ein paar Sätze zum Thema zu überlegen. Im Anschluss sollte reihum jeder Schüler einen informativen Satz zum Thema *Chile* äußern, welcher die bereits gehörten Informationen der anderen Schüler nicht wiederholen durfte.

Derartige Auflistungen können dem Lehrer und den Schülern somit dabei helfen, das bei der Behandlung des jeweiligen Themengebiets von den Schülern zu beherrschende Vokabular zu ermitteln. Unter Umständen kann es jedoch von Vorteil sein, mit getaggten Texten zu arbeiten, da die Lexeme in diesem Fall aufgrund der Angabe der Wortart genauer den Lemmata zugeordnet werden können.

Neben der Verarbeitung bloßer Textdateien lässt *AntConc* auch die Arbeit mit getaggten Texten zu.¹²⁷³ Beim *tagging* (siehe Kapitel 1.4.5.) werden die *tokens* in den Texten mit einem Wortartenlabel, einem sogenannten *tag*, versehen.¹²⁷⁴ Diese um Informationen zur Wortart erweiterten Dateien ermöglichen Suchanfragen nach bestimmten Wortarten wie Verben oder Substantiven. Es kann somit beispielsweise

¹²⁷² Bei den Texten, die für die Unterrichtseinheit zum Thema Chile verwendet wurden, handelt es sich um folgende Quelle: Arriagada Espinoza, Melanie (2012): *¡Adelante! Nivel avanzado. Schülerbuch*. Stuttgart: Klett, S. 25-31, 35.

¹²⁷³ Für die Erstellung der folgenden Informationen zum Umgang mit getaggten Texten wurde auf das Tutorial 11 der Version *AntConc 3.4.0* von Laurence Anthony zurückgegriffen: Vgl. Anthony, Laurence (2014): *AntConc 3.4.0 — Tutorial 11: Working with tagged data*. Online Publikation: <https://www.youtube.com/watch?v=gkKna-ka9zw> (25.02.2016).

¹²⁷⁴ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 190.

mittels des Suchausdrucks *VL** nach allen im Korpus vorkommenden lexikalischen Verben (Vollverben, unter Ausschluss der Hilfsverben) gesucht werden.

Die bisher im Programm *AntConc* verwendeten Textdateien können relativ einfach mit dem Programm *TagAnt* von Laurence Anthony in getaggte Texte verwandelt werden. Der *tagger TagAnt* ist, ebenso wie das Programm *AntConc*, kostenlos im Internet herunterzuladen.¹²⁷⁵ Die von *TagAnt* verwendeten *tags* (also das *tagset*), entsprechen denjenigen des Programms *TreeTagger*.¹²⁷⁶

Die Arbeit mit getaggten Texten bietet den Vorteil, dass ambigüe Lexeme, die vom Programm *AntConc* ohne weitere Informationen nicht eindeutig zugeordnet werden können, im Zweifelsfall durch die *tags* kategorisiert werden. Dadurch wird folglich ein genaueres Ergebnis erreicht. Im Fall der *Keyword List* des *Argentina*-Korpus macht es jedoch keinen Unterschied, ob mit getaggten Texten gearbeitet wird oder nicht, da die bereits erwähnte, im Hinblick auf Homonymie nicht eindeutige Form *para* als Präposition innerhalb der *Stoplist* ohnehin ausgeschlossen wird und die Form *trabajo* lediglich als Substantiv und nicht als Verbform in den Texten vorkommt. Für Untersuchungen zur Verwendung von Präpositionen wie *por* und *para* oder für ein anderes Korpus kann die Verwendung getaggtter Texte allerdings sehr aufschlussreich sein.

Es muss jedoch hinzugefügt werden, dass auch im Fall des automatischen *taggers TagAnt* und nicht nur bei umfassenden Referenzkorpora Lexeme falsch kategorisiert werden, wie bereits in Kapitel 1.4.5. erwähnt. Werden beispielsweise die Texte des *Adelante*-Korpus mithilfe des Programms *TagAnt* getaggt und in das Programm *AntConc* geladen, so müsste es theoretisch möglich sein, die Erscheinungen von *trabajo* als Substantiv einerseits und als Verbform andererseits unterscheiden zu können. Lässt man sich allerdings die Konkordanzen anzeigen, so ergibt sich ein anderes Bild, denn die Wortform *trabajo* wird fälschlicherweise ausschließlich als Substantiv und nicht als Verbform kategorisiert.

Somit wird klar, dass anhand des Programms *AntConc* zwar mit getaggten Texten gearbeitet werden kann und jegliche Texte mit dem Programm *TagAnt* getaggt werden können, falsche Zuordnungen der automatischen Annotation allerdings auch hier

¹²⁷⁵ Vgl. Anthony, Laurence (2015): *TagAnt (Version 1.2.0) [Macintosh OS X]*. Tokyo, Japan: Waseda University. Available from <http://www.laurenceanthony.net/> (23.05.2016).

¹²⁷⁶ Vgl. Schmid, Helmut: *TreeTagger – Spanish Tagset*. Online Publikation: <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/spanish-tagset.txt> (24.05.2016).

vorkommen und der Nutzer daher im Zweifelsfall die Konkordanzen zurate ziehen sollte. Für die Nutzung eines Korpus zu schulischen Zwecken sei jedoch gesagt, dass die grundlegenden Funktionen eines eigenen Korpus ebenso gut, wenn nicht sogar noch besser, anhand eines nicht getaggtten Korpus genutzt werden können, da die Arbeit mit bloßen Texten übersichtlicher ist.

4.7.3. Anwendungsbeispiele für die Nutzung eines selbst erstellten Korpus

Im Folgenden werden einige Anwendungsgebiete für die soeben beschriebenen Funktionen des Programms *AntConc* vorgestellt, um weitere Nutzungsmöglichkeiten für den Fremdsprachenunterricht aufzuzeigen.

4.7.3.1. Erstellung eigener Stoplisten im Programm *AntConc*

In Kapitel 4.7.2. wurde erklärt, wie man eine bereits existierende Stopliste im Programm *AntConc* verwendet. Derartige Stoplisten können jedoch auch selbst für ein Korpus erstellt werden, wie im Folgenden zu sehen ist. Bisher bestand die verwendete Stopliste aus einer vorgefertigten, aus dem Internet heruntergeladenen Auflistung frequenter Lexeme, welche um einige Funktionswörter erweitert wurde. Die Verwendung dieser Stopliste hat den Vorteil, dass eben diese ohnehin bekannten, da generell sehr häufig vorkommenden Lexeme in den Ergebnissen ausgeblendet werden. Um nun eine Übersicht mit neu zu erlernendem Vokabular zu einem Themengebiet zu ermitteln, erscheint es sinnvoll, nicht nur die besonders frequenten Lexeme, sondern, sofern möglich, generell das bereits bekannte Vokabular auszublenden. Zu diesem Zweck kann beispielsweise eine Stopliste aus dem *Adelante*-Korpus erstellt werden. Angenommen, die Lehrkraft möchte nach der Bearbeitung des Themengebiets *Cataluña* (*¡Adelante! Nivel intermedio - Unidad 3*) landeskundliche Texte zu den Themen Fußball, Architektur und Landschaft Kataloniens mit der Klasse lesen. Es sind also bisher das Lehrbuch *¡Adelante! Nivel elemental* komplett und *¡Adelante! Nivel intermedio* bis zu Kapitel 3 bearbeitet worden. Anhand einer Stopliste könnte nun ermittelt werden, welches Vokabular eines neuen Texts oder eines neuen Themengebiets noch unbekannt ist.

Zu diesem Zweck wird aus den bis dahin gelesenen Texten mithilfe der Programme *AntConc* und *Excel* eine *Word List* erstellt, welche alle bis dato gelernten Lexeme enthält. Diese *Word List* kann daraufhin als Stopliste in das Programm *AntConc* geladen werden, um sie mit neuen Texten abzugleichen. Bei der Erstellung einer *Word List* zu einem Text

über den *FC Barcelona* werden unter Verwendung der Stopliste beispielsweise die Lexeme *FC*, *campeones*, *camp*, *liga*, *líder*, *Nou* und *copa* als neues Vokabular ermittelt. Neben diesem Spezialwortschatz zum Thema Fußball finden sich allerdings auch Internationalismen wie *democracia* oder *capacidad* sowie generell neues Vokabular, das nicht unmittelbar mit dem Themengebiet verbunden ist, in der *Word List*. Diese Auflistung kann den Schülern als Vorbereitung auf die Arbeit mit dem neuen Text an die Hand gegeben werden oder die Basis für eine vorherige Brainstorming-Phase bilden. Die Lehrkraft könnte die Schüler fragen, welche Lexeme ihnen bereits bekannt sind, welche sie sich herleiten können und welche sie noch nicht kennen. Auf diese Weise können die Schüler vorab unbekanntes Vokabular klären, um den Text besser zu verstehen. Andererseits kann eine derartige Liste auch der Nachbereitung zu Hause dienen.

Auf die gleiche Weise kann mit anderen neuen Texten, beispielsweise zur Architektur von Antoni Gaudí, der *Costa Brava* oder dem *Museo Dalí* verfahren werden. Derartige *Word Lists* können zudem nicht nur zu einzelnen Texten, sondern auch zu einem themenbezogenen Korpus erstellt werden. Ein Beispiel ist die Erstellung einer *Word List* zum Thema Chile in der 12. Klasse, also dem 3. Lernjahr im Fach Spanisch als spätbeginnende Fremdsprache, welches bereits in Kapitel 4.7.2. erwähnt wurde. In besagtem Kurs wurde das Thema Chile als zweites Themengebiet im Halbjahr 12/2, nach der Bearbeitung von Argentinien, durchgenommen. Daher stellen alle Texte der Lehrwerke *¡Adelante! Nivel elemental* und *¡Adelante! Nivel intermedio* sowie die Dateien zum Thema Argentinien die Grundlage für die Stopliste dar. Wenn man nun mithilfe des Programms *AntConc* die Texte des Themengebiets Chile mit der Stopliste abgleicht, werden unter anderem *mapuche*, *minería*, *socialista*, *agraria* oder *cordillerana* aufgelistet. Die vollständige Liste der laut *Chile*-Korpus neuen Lexeme findet sich in Arbeitsblatt 28 *Chile – Vocabulario nuevo* (Anhang 37), wobei Namen, beispielsweise von Personen und Orten, gelöscht wurden. In der Auflistung befindet sich neben bis dato unbekannten Lexemen auch bereits bekanntes Vokabular, wie Verben, bei denen lediglich diese Form vorher nicht in den Schulbuchtexten behandelt wurde, wie beispielsweise die dritte Person Singular des *pretérito indefinido* von *haber* – *hubo*. Diese Verbform wurde zwar im Rahmen der Grammatikeinheit über das *pretérito indefinido* vermittelt, sie ist aber nicht in den Texten erschienen. Außerdem werden manche Lemmata doppelt aufgezählt, wenn zum Beispiel zwei Formen neu hinzukommen, wie

bei *invirtieron* und *invirti6*. Man könnte diese doppelte Aufzählung zwar durch die Arbeit mit einer Lemmaliste umgehen (siehe Kapitel 4.7.2.), dadurch würde aber stets die Grundform der Lexeme in der Liste erscheinen, was bei Verben wie *haber*, *hacer* und *ser* unsinnig scheint, wenn man davon ausgeht, dass die bislang in Lehrbuchtexten nicht erschienene Verbform, also in diesem Fall *hubo* oder *invirti6*, noch einmal wiederholt werden sollte.

Diese Liste neuer Vokabeln für das Themengebiet Chile kann die Lehrkraft den Schülern ebenso wie beim Thema Argentinien vor, während oder nach der Behandlung des Themas an die Hand geben. Sie kann allerdings auch lediglich vom Lehrer genutzt werden, um neues Vokabular für Tests oder auch für die Semantisierung in den einzelnen Unterrichtsstunden zu ermitteln, wobei hier auch *Word Lists* zu einzelnen Texten sinnvoll erscheinen.

Die *Chile Word List* ähnelt der bereits beschriebenen *Keyword List* mit den wichtigsten Lexemen des Themengebiets Chile (siehe Kapitel 4.7.2.). Je nachdem wofür der Lehrer die Vokabelliste verwenden möchte, erscheint die eine oder die andere Liste sinnvoller. Möchte die Lehrkraft den Schülern vorab eine Liste mit den im folgenden Themengebiet am häufigsten vorkommenden Lexemen geben, so ist die erste Liste, die *Keyword List*, sicherlich die richtige Wahl. Anhand dieser Liste kann die Lehrkraft den Schülern zudem zeigen, dass ihnen viele der dort erwähnten Lexeme bereits bekannt sind, und sie kann somit auch für Individualisierung und Differenzierung sorgen. Außerdem kann sie als Anregung für ein abschließendes Gespräch zum Thema dienen, da auch frequente Namen enthalten sind. Wird jedoch eine Liste aller bisher unbekannten Lexeme, die für die Bearbeitung der Texte zum Thema Chile relevant sind, benötigt, beispielsweise als Wortschatzliste, so wird die Wahl auf die zuletzt erstellte *Chile Word List* fallen, welche im Fall von Textsammlungen die Funktion des Vokabelteils am Ende des Schulbuchs übernehmen kann.

4.7.3.2. Erstellung von Grammatikübungen anhand bereits bekannter Texte

Ein Beispiel für die Erarbeitung von Grammatikübungen ist die Erstellung eines Arbeitsblatts zur Unterscheidung der Präpositionen *por* und *para*. Auch fortgeschrittene Lerner haben bisweilen Probleme mit der Verwendung dieser beiden Präpositionen, daher ist ein Arbeitsblatt zu dieser Grammatikeinheit im Themengebiet *Argentina* denkbar, welches, wie bereits erwähnt, im dritten Lernjahr angesetzt ist. So können die Sätze aus

den bereits im Unterricht gelesenen Texten als Beispielsätze dienen, anhand derer die Lehrkraft ein Arbeitsblatt erstellt. Unter den Konkordanzen müssen zunächst passende Beispielsätze ausgesucht werden. Im Programm *AntConc* können verschiedene Konkordanzen ausgewählt, aber auch gelöscht werden. Die Konkordanzen können entweder als bloße Ergebnisliste in Form einer *txt*-Datei exportiert oder aber zuvor bearbeitet werden. Im vorliegenden Fall soll eine Lückenübung erstellt werden, weshalb die Option *Hide search term in KWIC display* ausgewählt wird. Diese ermöglicht es, dass die Konkordanzen nicht mit dem Suchausdruck, sondern mit Platzhaltern dargestellt werden. Mittels dieser Option können auf relativ einfache Weise Aufgaben zu den Bereichen Wortschatz oder Grammatik entwickelt werden. In diesem Fall werden die Präpositionen *por* und *para* durch Platzhalter ersetzt, um ein Arbeitsblatt zu erstellen, mithilfe dessen die Schüler den Gebrauch von *por* und *para* im Themengebiet Argentinien wiederholen sollen. Das mit dem *Argentina*-Korpus erstellte Arbeitsblatt für das dritte Lernjahr Spanisch befindet sich in Anhang 38 (siehe Arbeitsblatt 29 *Las preposiciones por y para*). Es konnte von den Schülern der 12. Jahrgangsstufe problemlos bearbeitet werden und sie schätzten die Vertiefung dieses Grammatikkapitels anhand ihnen bereits bekannter Texte sehr, da sie keinerlei Wortschatz nachschlagen mussten.

4.7.3.3. Erstellung eines Korpus mit Schülern

Anhand des Programms *AntConc* kann neben der Erstellung von Unterrichtsmaterialien ein fachspezifisches oder fächerübergreifendes Projekt durchgeführt werden – beispielsweise in Zusammenarbeit mit dem Fach Informatik. Es könnte ein eigenes klassenspezifisches oder progressives Korpus (siehe Monitorkorpora in Kapitel 2.7.) in einem Workshop oder einem fortlaufenden unterrichtsbegleitenden Projekt erstellt werden. Das Thema kann zu Beginn des Schuljahrs gemeinsam festgelegt werden, indem die Schüler beispielsweise das Schulbuch durchsehen und das Thema bestimmen, welches sie am meisten interessiert, oder aber der Lehrer gibt eine Auswahl vor. Durch die aktive Mitarbeit an diesem Korpus werden die Schüler motiviert, sind interessiert am Korpus und folglich auch an der späteren Korpusarbeit. In dieses Korpus können nicht nur von den Schülern selbst ausgewählte, reale Texte aus dem Internet oder anderen digitalen Medien integriert werden, sondern auch die Lehrbuchtexte, welche im bis dato durchgeführten Unterricht verwendet wurden, damit die Umstellung auf die Arbeit mit Korpora und Internettexen nicht zu abrupt ist. Die Schüler haben also

gleichermaßen bekannte wie unbekannte Texte vorliegen und können sich so allmählich an die Arbeit mit nicht-didaktisierten Texten gewöhnen. Die abgerufenen Konkordanzen zeigen unter diesen Umständen ebenfalls teilweise bekannte, dem Lehrbuchstil entsprechende Textausschnitte und vermitteln dem Schüler so ein Gefühl der Vertrautheit.

Eine andere Möglichkeit besteht darin, das Thema eines gemeinsam mit der Klasse erstellten Korpus je nach Interessenlage auszuwählen. So bietet es sich an, in einem Jahr, in welchem eine Fußballwelt- oder Europameisterschaft stattfindet, in einer an Fußball interessierten Klasse ein Korpus aus Texten zum Thema Fußball zusammenzustellen. In dieses Korpus werden wohl zu einem großen Teil aktuelle Zeitungstexte geladen werden, es können aber auch allgemeine Informationstexte zu Spielregeln, über die jeweilige Meisterschaft oder bestimmte Spieler entweder aus dem Internet oder anderen digitalen Quellen mit aufgenommen werden. Durch die gemeinsame Erstellung des Fußball-Korpus wählen die Schüler zunächst geeignete und für sie interessante Texte aus, sammeln diese im Korpus, bestimmen Schlüsselwörter des Themengebiets und erarbeiten Vokabellisten. Zudem kann die Lehrkraft anhand dieses Korpus Besonderheiten der Zeitungssprache veranschaulichen sowie Arbeitsblätter zu relevanten grammatischen Strukturen erstellen. Da die Schüler an der Auswahl des Themas sowie der Texte aktiv beteiligt sind, liegt ein gewisses Grundinteresse vor, was die Lehrkraft für die Bearbeitung eben dieser Grammatikeinheiten nutzen kann. Durch die intensive Auseinandersetzung mit den Zeitungstexten ist es außerdem vorstellbar, die Schüler gegen Ende der Unterrichtssequenz einen Zeitungstext zu einem aktuellen Spiel, einem Vorfall während der Meisterschaft oder über einen bestimmten Spieler beziehungsweise eine Mannschaft schreiben zu lassen. Die Schüler können somit ihr Wissen über den Aufbau sowie die typischen Strukturen eines Zeitungstexts für die Erstellung eines eigenen Artikels nutzen, und die Lehrkraft kann daran den Erfolg der Aneignung des Spezialwortschatzes sowie der gattungsspezifischen Eigenheiten überprüfen.

Alternativ kann eine ähnliche Vorgehensweise zum Thema ‚Märchen‘ gewählt werden. Auch hier können verschiedene spanische Märchen gesammelt, anhand des entstandenen Korpus Vokabular, Kollokationen, grammatische Strukturen sowie Besonderheiten dieser Gattung erforscht und gegen Ende selbst, unter Beachtung der zuvor untersuchten Gattungsmerkmale, ein Märchen verfasst werden. Dieses Thema

eignet sich zudem besonders für die Veranschaulichung der Vergangenheitszeiten und deren Verwendung, wodurch diese für Schüler in der Regel problematische Grammatikeinheit ausgebaut oder wiederholt werden kann. Ebenso kann allerdings ein Korpus lediglich zu einem bestimmten Thema, wie einem spanischsprachigen Land, einer Stadt, Sehenswürdigkeiten, Essen oder Naturkatastrophen erstellt werden, ohne auf ein spezielles Genre zu achten.

Des Weiteren ist es durchaus denkbar, in einem Oberstufenkurs Spanisch als dritte Fremdsprache, also im 4. oder 5. Lernjahr, die Schüler anhand eines Korpus auf die in der schriftlichen Abiturprüfung gängigen Themen vorzubereiten. Die von der Lehrkraft ausgewählten Texte einer bestimmten Textgattung können in einem Korpus verhältnismäßig schnell nach bestimmten Lemmata, grammatischen Strukturen oder Kollokationen durchsucht werden, welche sich die Schüler durch ihre aktive Beteiligung besser einprägen. Dieses Wissen können die Schüler wiederum für ihre eigene Textproduktion verwenden. Anhand der Frequenzliste kann außerdem ermittelt werden, welche Lemmata auf jeden Fall gelernt werden sollten.

Wie in diesem Exkurs zu sehen war, können in das Programm *AntConc* alle im Spanischunterricht besprochenen Texte hochgeladen werden, womit der Lehrer ein speziell auf den jeweiligen Unterricht zugeschnittenes Korpus zur Verfügung hat. Man kann anhand dieses Korpus genau überprüfen, welches Vokabular bereits im Unterricht verwendet wurde und daher in einer Schulaufgabe oder auf Arbeitsblättern als bekannt vorauszusetzen ist. Im Gegensatz zu den digitalen Vokabelübersichten gängiger Lehrwerke bietet ein auf diese Weise erstelltes Korpus die Möglichkeit, auch vom Lehrer ausgewählte und im Unterricht besprochene Zeitungsartikel, Texte aus anderen Lehrwerken, Informationsbroschüren oder Ähnliches zu integrieren und alle Texte zusammen zu durchsuchen.

Die Erstellung eines eigenen Korpus ist sicherlich mit einem gewissen Arbeitspensum verbunden, was viele Lehrkräfte davon abhalten wird, doch wie im Verlauf dieses Exkurses gesehen werden konnte, ist es auch für eine Lehrkraft ohne spezielle IT-Kenntnisse anhand eines Programms wie *AntConc* durchaus möglich, ein derartiges Korpus zu entwickeln. Mithilfe eines eigenen Korpus aus allen im Unterricht behandelten Texten kann die Lehrkraft unter anderem für die Erstellung von Tests und

4. Nutzung von Korpora in der fremdsprachlichen Unterrichtspraxis

weiteren Unterrichtsmaterialien Fragen zum Vokabular klären, Wortschatzlisten für verschiedenste Zwecke zusammenstellen sowie Arbeitsblätter zu bestimmten Grammatikeinheiten erarbeiten, welche Beispielsätze aus dem jeweiligen Themenbereich umfassen. Sofern auch die Schüler Zugang zum Korpus haben, können sie sich Anwendungsbeispiele für Vokabular und Grammatikeinheiten, deren Verwendungsweise ihnen noch unklar ist, anzeigen lassen, wobei es sich um Sätze aus ihnen bereits bekannten Texten handelt und ihnen das Verständnis dadurch erleichtert wird. Generell wird die Erstellung und Nutzung eines derartigen pädagogischen Korpus allerdings in der Mehrheit der Fälle den Lehrkräften vorbehalten bleiben, da hinsichtlich der aktuellen technischen Ausstattung eine umfassende Korpusnutzung im und vor allem außerhalb des Unterrichts schwer zu realisieren sein wird. Dennoch kann die Lehrkraft den Schülern Auflistungen von Anwendungsbeispielen für Wortschatz und Grammatik im Unterricht vorlegen oder für die häusliche Vorbereitung auf Tests an die Hand geben, um ihnen die jeweiligen Lexeme oder Grammatikeinheiten anhand bekannter Texte zu veranschaulichen.

5. Schlussfolgerungen: Möglichkeiten und Grenzen der Verwendung von Korpora für den Fremdsprachenunterricht

Im Verlauf dieser Arbeit konnte gezeigt werden, dass Korpora durchaus sinnvoll und gewinnbringend in den fremdsprachlichen Unterricht integriert werden können und nicht *per se* zu komplex für die Nutzung durch Lehrer und Schüler im schulischen Kontext sind. Bei der Verwendung von Korpora im Unterricht bedarf es einer durchdachten und vorausschauenden Planung sowie einer intensiven Einführung der Schüler in die Funktionsweise derselben, doch die Arbeit mit Korpora bietet einige positive Aspekte, welche die auf die Einarbeitung verwendete Zeit wieder aufwiegen. Dennoch gibt es nicht nur Vorteile, sondern es werden von verschiedenen Wissenschaftlern auch Bedenken gegenüber der Nutzung von Korpora im Unterricht geäußert, weshalb im Folgenden auf beide Seiten noch einmal eingegangen werden soll. Zunächst soll jedoch ein kurzer Blick auf einige Vorteile und Probleme der Korpuslinguistik im Allgemeinen geworfen werden, da diese ebenso für die didaktische Verwendung von Korpora relevant sind.

Lemnitzer/Zinsmeister sehen als einen Vorteil der Korpuslinguistik, dass aufgrund der Datenbasis, die die Korpuslinguistik bereitstellt, nicht nur die Struktur einer Sprache, sondern auch deren Verwendung untersucht werden kann.¹²⁷⁷ Grundsätzlich sind durch Introspektion gewonnene Daten subjektiv, schwierig zu verifizieren und artifiziell. Zudem können keinerlei Aussagen zur Frequenz von Lexemen gemacht werden. Die umfassenden Sammlungen realer Texte erlauben hingegen den Zugang zu natürlichen Sprachdaten, auch unterschiedlicher Varietäten. Große Textmengen können auf schnelle Art und Weise und mit hoher Verlässlichkeit durchsucht werden. Man kann sagen, dass die sprachwissenschaftliche Forschung anhand der Korpuslinguistik technologisiert wurde.¹²⁷⁸

Die computergestützte Technologie kann zudem in allen Bereichen der Sprache verwendet werden: Phonetik, Morphologie, Syntax, Lexikologie, Pragmatik, Semantik sowie auf Ebene des Diskurses. Forscher können bereits bestehende Korpora für ihre

¹²⁷⁷ Vgl. Lemnitzer/ Zinsmeister (2010): op. cit., S. 10.

¹²⁷⁸ Vgl. Parodi (2010): op. cit., S. 167.

5. Schlussfolgerungen: Möglichkeiten und Grenzen der Verwendung von Korpora für den Fremdsprachenunterricht

eigenen Untersuchungen nutzen, wodurch sie sich viel Arbeit, Zeit und Kosten ersparen, die durch die Erstellung eines eigenen Korpus entstanden wären.¹²⁷⁹

Durch Korpora können Hypothesen anhand einer großen Anzahl von Belegen objektiv verifiziert sowie quantitative Daten erhoben werden.¹²⁸⁰ Sie dienen der Verifizierung und Falsifizierung von Hypothesen, weshalb sie von der theoretisch arbeitenden Linguistik nicht herabgesetzt, sondern als Ergänzung und Korrektiv ihrer aus der Intuition gewonnen hypothetischen Modelle betrachtet werden.¹²⁸¹ Kennedy schreibt, dass Wissenschaftler verschiedener theoretischer Überzeugungen heutzutage Korpora zur Überprüfung ihrer Theorien verwenden.¹²⁸²

Kennedy legt zudem dar, dass es falsch wäre, anzunehmen, dass die Korpuslinguistik lediglich ein schnellerer Weg zur Beschreibung einer Sprache sei oder dass es sich dabei nur um linguistische Belege handle. Die Analyse eines Korpus mittels gängiger korpuslinguistischer Software enthülle häufig Fakten, an deren Erforschung man zuvor nie gedacht hätte.¹²⁸³ Beispielsweise führte Altenbergs Studie zum Thema „Amplifier collocations in spoken English“ zu weiteren Fragen hinsichtlich semantischer Klassen von verstärkenden Adverbien wie *perfectly* oder *awfully*, welche ohne die Belege eines Korpus nicht hätten gestellt werden können. So fand er heraus, dass häufige verstärkende Adverbien wie *quite* dazu neigen, mit nicht skalierbaren Lexemen eine Kollokation einzugehen (*quite obviously*), während *absolutely* mehr als andere verstärkende Adverbien dazu tendiert, Kollokationen mit Negationspartikeln zu bilden (*absolutely not*).¹²⁸⁴ Außerdem werde durch die Korpuslinguistik der Fokus vielmehr auf die Frage danach gelenkt, welche Lexeme, Strukturen oder Verwendungsweisen in einer Sprache wahrscheinlich seien, als danach, welche überhaupt möglich seien.¹²⁸⁵

Buckingham weist darauf hin, dass ein Korpus viele Vorteile für die angewandte Linguistik mit sich bringt. So bereichert es unser Verständnis der relativen Häufigkeit bestimmter Ausdrücke, kontextualisiert diese zugleich und vereinfacht damit die

¹²⁷⁹ Vgl. Parodi (2010): op. cit., S. 167.

¹²⁸⁰ Vgl. Parodi (2010): op. cit., S. 167.

¹²⁸¹ Vgl. Klöden (2002): op. cit., S. 12-13.

¹²⁸² Vgl. Kennedy, Graeme (1998): *An introduction to corpus linguistics*. Harlow: Longman, S. 8.

¹²⁸³ Vgl. Kennedy (1998): op. cit., S. 8-9.

¹²⁸⁴ Vgl. Altenberg, Bengt (1991): „Amplifier collocations in spoken English“. In: Johansson, S./ Stenström, A.-B. (Hrsg.) (1991): *English Computer Corpora. Selected Papers and Research Guide*. Berlin/ New York: Mouton de Gruyter, S. 127-148.

¹²⁸⁵ Vgl. Kennedy (1998): op. cit., S. 8-9.

Formulierung von Verallgemeinerungen in Bezug auf semantische und pragmatische Aspekte sowie morphosyntaktische Variationen. Außerdem ist es nicht nur eine Quelle für reale und typische, sondern auch für außergewöhnliche Beispiele, welche unter anderem für lexikografische Zielsetzungen verwendet werden können.¹²⁸⁶

Stefanowitsch führt zudem aus, dass sich die Sprachwissenschaft mit einer Verstärkung der Arbeit in der Korpuslinguistik, insbesondere in der quantitativen Korpuslinguistik, von einer reinen Geisteswissenschaft zu einer „harten Wissenschaft“ nach dem Vorbild der Naturwissenschaften entwickeln kann, ohne deren geisteswissenschaftliche Herkunft jedoch zu verleugnen.¹²⁸⁷

Er spricht allerdings auch allgemeine Probleme der derzeitigen Korpuslinguistik an, welche ebenso bei der Arbeit mit Korpora im Unterricht bedacht werden müssen. Stefanowitsch schreibt, dass der Vorteil großer Korpora auch ein Nachteil ist. Die enorme Größe ermöglicht zwar eine Arbeit mit vielen sprachlichen Beispielen und einen hohen Grad an Repräsentativität, sie führt aber auch dazu, dass die Annotationen dieser Korpora nicht manuell durchgeführt werden können, sondern von statistischen sprachverarbeitenden Systemen vorgenommen werden müssen. Generell schätzt man, dass bei dieser maschinellen Annotation ungefähr 3% der Lexeme falsch kategorisiert werden. Allerdings sieht Stefanowitsch darin kein wirkliches Problem. Seiner Meinung nach ist die Datenmenge solcher Korpora so groß, dass die geringe Fehlerquote bei der Kategorisierung die Ergebnisse insgesamt nicht infrage stellen dürfte. Außerdem, so Stefanowitsch, werden auch bei einer sorgfältigen manuellen Annotation Kodierungsfehler vorkommen.¹²⁸⁸ Diese Kritik ist jedoch durchaus angebracht, denn wie im vierten Kapitel zu sehen war, führt die automatische Annotation zu falschen Kategorisierungen in Bezug auf die Wortart, was bei der Erstellung von Wortfamilien bisweilen zu Schwierigkeiten führen kann.

Einen weiteren Nachteil sieht Stefanowitsch darin, dass Korpuslinguisten trotz ihrer empirischen Vorgehensweise bei Suchen im Korpus immer wieder intuitiv Entscheidungen treffen müssen. So ist schon die Einteilung der Lexeme in verschiedene Wortarten nicht bei allen Sprachwissenschaftlern die gleiche. Zudem müssen die

¹²⁸⁶ Vgl. Buckingham (2009): op. cit., S. 45.

¹²⁸⁷ Stefanowitsch (2005): op. cit., S. 160.

¹²⁸⁸ Vgl. Stefanowitsch (2005): op. cit., S. 157-158.

5. Schlussfolgerungen: Möglichkeiten und Grenzen der Verwendung von Korpora für den Fremdsprachenunterricht

Ergebnisse einer Suchanfrage immer auch interpretiert werden und eine Interpretation wird zwangsläufig von persönlichen Einstellungen und dem Wissen des Suchenden beeinflusst.¹²⁸⁹ Klöden sieht dies ähnlich. Obgleich die Entwicklung der maschinellen Korpuserstellung bereits enorme Fortschritte gemacht hat, so Klöden, ist eine letzte manuelle Auswertung unabdingbar, da reine Frequenzauszählungen weitgehend wertlos sind. So kann die Verwendung zweier Tempora zahlenmäßig durchaus mithilfe eines Korpus ermittelt werden, doch die Frage danach, ob die beiden Zeiten tatsächlich austauschbar sind, kann nur durch eine inhaltliche Auswertung geklärt werden.¹²⁹⁰

Außerdem gibt Klöden an, dass Korpora *per definitionem* begrenzt sind und Untersuchungsergebnisse unterschiedlicher Korpora daher nicht unbedingt miteinander verglichen werden können. Die Begrenztheit führt auch dazu, dass Korpora im Hinblick auf eine gesprochene Sprache einen statischen Eindruck einer dynamischen Realität vermitteln. Des Weiteren werden je nach der Thematik der darin befindlichen Texte bestimmte lexikalische Einheiten über- oder unterrepräsentiert, und manche erscheinen gar nicht im Korpus.¹²⁹¹

Stubbs legt zudem dar, es würde häufig kritisiert, dass die Methoden der Korpusarbeit einige sprachliche Aspekte, wie beispielsweise Kollokationen, zu sehr betonten. Diese Kritik könne seiner Meinung nach durchaus gerechtfertigt sein, doch die Beobachtungswerkzeuge seien eben dazu da, etwas Bestimmtes hervorzuheben. Die Errungenschaften der Naturwissenschaften beispielsweise seien nur möglich gewesen durch die Entwicklung von Beobachtungswerkzeugen wie dem Teleskop, dem Mikroskop und der Kamera, da sie Muster enthüllt hätten, welche für das bloße Auge nicht sichtbar gewesen seien. Ein großes Korpus mit der entsprechenden Software könne unter diesem Aspekt durchaus mit derartigen Werkzeugen verglichen werden, da es sprachliche Muster enthüllen könne, die mit dem bloßen Auge eben nicht zu erkennen seien. Bei einem Mikroskop oder einem Teleskop beschwere sich jedoch niemand darüber, dass das Eine klitzekleine und das Andere weit entfernte Dinge betone.¹²⁹² Auch

¹²⁸⁹ Vgl. Stefanowitsch (2005): op. cit., S. 158.

¹²⁹⁰ Vgl. Klöden (2002): op. cit., S. 10.

¹²⁹¹ Vgl. Klöden (2002): op. cit., S. 10-12.

¹²⁹² Vgl. Stubbs, Michael (2000): 'Society, education and language: the last 2,000 (and the next 20?) years of language teaching'. In: Trappes-Lomax, H. (Hrsg.) (2000): *Change and Continuity in Applied Linguistics*. Clevedon: BAAL and Multilingual Matters, S. 15-34. Online Publikation: <https://www.univ-trier.de/fileadmin/fb2/ANG/Linguistik/Stubbs/stubbs-2000-soc-ed-lg.pdf> (19.07.2017), S. 12.

Hunston führt aus, Korpora sollten nicht für etwas kritisiert werden, was sie nicht sind. Korpora und dazugehörige Programme könnten beispielsweise Konkordanzen präsentieren und Häufigkeiten ermitteln. Andere Forschungsmethoden, welche mit ihnen nicht möglich seien, müssten auf andere Weise angegangen werden.¹²⁹³

Andererseits vertritt Stefanowitsch die Meinung, dass Fragen, auf welche die Korpuslinguistik Antworten liefert, ohne die Existenz großer Korpora gar nicht hätten gestellt werden können. Angaben zur Häufigkeit von Lexemen oder zu deren gemeinsamem Auftreten wären ohne diese Korpora und die moderne Technik der Korpuslinguistik nicht denkbar. Nicht nur Assoziationen zwischen Lexemen, sondern auch solche zwischen bestimmten Lexemen und grammatischen Strukturen lassen sich mithilfe großer elektronischer Korpora ermitteln. Angaben darüber könnten ohne Korpora nur vermutet, aber nie empirisch begründet werden.¹²⁹⁴ Hunston sieht dies ähnlich und legt dar, dass die Korpusarbeit sich auf Sprachtheorien und Sprachbeschreibungen auswirkt. Durch Korpora können nicht nur die traditionellen Kategorien ausgezählt, sondern es können auch Kategorien und Phänomene erforscht werden, welche bis dahin noch nicht bemerkt worden sind. Sie schreibt überdies, dass Korpora die Erforschung und Nutzung von Sprachen revolutioniert haben und daran nicht zuletzt die verbesserte Verfügbarkeit von Computern mitgewirkt hat.¹²⁹⁵ Die Erkenntnisse, welche die Korpuslinguistik liefert, können wiederum für den Fremdsprachenunterricht und die darin vermittelte Sicht auf die Sprache sowie für die Entwicklung von Unterrichtsmaterialien genutzt werden. Generell ist der Erkenntnisgewinn, welchen die Arbeit mit Korpora für die Sprachwissenschaft im Allgemeinen mit sich bringt, ebenso dem Fremdsprachenunterricht dienlich. Zudem können die Korpora selbst auch im und für den Unterricht genutzt werden, weshalb im Folgenden die Möglichkeiten und Grenzen der didaktischen Nutzung von Korpora beleuchtet werden sollen.

Korpora dienen der Lehrkraft als Quelle für Lehrmaterialien. Sie können den Schülern eine sinnvolle Auswahl und Menge realer Beispiele bieten. Wie im dritten und vierten Kapitel zu sehen war, können Korpora allerdings nicht nur von der Lehrkraft für

¹²⁹³ Vgl. Hunston (2002): op. cit., S. 20.

¹²⁹⁴ Vgl. Stefanowitsch (2005): op. cit., S. 155.

¹²⁹⁵ Vgl. Hunston (2002): op. cit., S. 1.

5. Schlussfolgerungen: Möglichkeiten und Grenzen der Verwendung von Korpora für den Fremdsprachenunterricht

die Unterrichtsvorbereitung verwendet werden, sondern Schüler können ein Korpus zur Bearbeitung eines Arbeitsblatts sowie für eigene Recherchen nutzen.

Die Arbeit mit Korpora ermöglicht damit eine weitgehende Individualisierung des Lernens. Die Schüler können in ihrer eigenen Geschwindigkeit im Korpus suchen, Konkordanzen auswählen, die ihnen relevant erscheinen, und diese analysieren. Sofern der Lehrer freie Aufgaben stellt, können die Schüler nach ihrem Willen mit dem Korpus arbeiten und später in der Klasse ihre Ergebnisse vergleichen. Bei Unterrichtseinheiten, in welchen jeder Schüler oder jede Gruppe von Schülern andere Aufgaben zu lösen hat, fühlen sich einzelne nicht durch die Geschwindigkeit der Klassenkameraden bedrängt, da sie ihr eigenes Spezialgebiet bearbeiten und ihre Ergebnisse am Ende der Unterrichtseinheit dem Rest der Klasse präsentieren. Auf diese Art und Weise kann durch die Arbeit mit Korpora der viel geforderten Binnendifferenzierung für den schulischen Unterricht Rechnung getragen werden.

Die Arbeit mit Korpora bietet zahlreiche Einsatzmöglichkeiten. Man kann mit Korpora strukturiert arbeiten, sofern ein vom Lehrer erstelltes Arbeitsblatt vorliegt und die Suchanfragen relativ genau vorgezeichnet sind. Es ist aber auch experimentelles Lernen möglich, wenn die Schüler ohne genaue Vorgaben im Korpus stöbern dürfen und selbst bestimmen, wonach sie suchen und worauf sie genauer eingehen. Diese Freiheit erfahren die Schüler zum Beispiel, wenn der Lehrer ihnen ein Problem zeigt und sie dieses mithilfe eines Korpus lösen sollen.

Des Weiteren können die Schüler nicht nur alleine Suchabfragen durchführen und Ergebnisse auswerten, sondern mit Mitschülern in kleinen Gruppen zusammenarbeiten sowie komplexere Suchen aufteilen. Sie können die Korpora nicht nur in der Schule verwenden, um die vom Lehrer gestellten Aufgaben zu lösen, sondern ihr Wissen auch zu Hause bei Textproduktionen anwenden und mithilfe der Korpora als Rechercheinstrument Entscheidungen über Wortwahl, Kollokationen und grammatische Strukturen treffen.

Korpora fördern damit selbstständiges und individuelles Lernen anhand realer Sprachdaten und bieten somit mindestens zwei Motivationsfaktoren für Schüler – selbstreguliertes Lernen und Ansporn durch den Umgang mit muttersprachlichem Material. Anhand der Texte und Konkordanzen der Korpora können Schüler ihre Sprachkenntnisse erweitern und Ideen für die eigene Sprachproduktion sammeln. Sie

verstehen Zusammenhänge zwischen Lexemen, wie es beispielsweise bei der Arbeit mit Kollokationen und Phraseologismen der Fall ist, und ermitteln, mit welcher Häufigkeit bestimmte Ausdrücke von Muttersprachlern verwendet werden. Nicht nur das Vokabular, sondern auch das Wissen über grammatische Strukturen oder regionale Varianten können mithilfe von Korpora erweitert werden. Außerdem liefern vor allem repräsentative Korpora viele Belege für ein gesuchtes Lexem oder eine Phrase und präsentieren diese auf anschauliche Weise in den Konkordanzen, also im Format *Key Word In Context* (siehe Kapitel 1.4.3.), wodurch Nutzer die Verwendung sowie den typischen Kontext des jeweiligen Ausdrucks auf einen Blick sehen und genauer untersuchen können. Zudem können durch die Vielzahl an Belegen charakteristische Merkmale gewisser Genres oder Varietäten teilweise erst erkannt werden. Wie zu sehen war, geht die Nutzung von Korpora somit einher mit der Realisierung wichtiger Forderungen des Fremdsprachenunterrichts: Es werden reale Texte oder Ausschnitte daraus im Unterricht verwendet, Korpora tragen zur kommunikativen Kompetenz bei und können den Lerner im selbstgesteuerten Lernen unterstützen.

Die selbstständige Suche in Korpora hat einen zusätzlichen Vorteil, denn wie Diana Laurillard es formuliert: “There is no question that the new technologies are attractive to students, and that they engage attention by the very fact of their interactive nature.”¹²⁹⁶ Es könnten also schon allein wegen des Interesses der Schüler an neuen Technologien hin und wieder Korpora zur selbstständigen Suche im Unterricht verwendet werden. Grundsätzlich ist die Nutzung von Korpora in der Schule schon allein daher zu fördern, weil die Schüler im Unterricht mit Korpora am Computer selbst aktiv werden. Die Suche nach Lexemen, Wortfamilien, Phraseologismen oder grammatischen Eigenheiten kann das Interesse der Schüler wecken und sie motivieren, da sie am Computer selbst tätig sind.

Dennoch muss eines klar sein: Technologie alleine führt nicht automatisch zu gutem und nachhaltigem Lernen. Die Technologie ist lediglich ein Hilfsmittel und muss als solches in einen sinnvollen und klar durchdachten pädagogischen Rahmen eingebettet werden, um zum gewünschten Ergebnis zu gelangen. Die Schüler sollten von der Lehrkraft gelenkt, aber nicht rigide geführt werden. Der Lehrer spielt dabei eine wichtige

¹²⁹⁶ Laurillard, Diana (1998): *How Can Learning Technologies Improve Learning?*. Online Publikation: <http://web.archive.org/web/20070322002729/http://www.law.warwick.ac.uk/ltj/3-2j.html> (21.07.2011).

5. Schlussfolgerungen: Möglichkeiten und Grenzen der Verwendung von Korpora für den Fremdsprachenunterricht

Rolle, da die Schüler, zumindest anfangs, nicht mit den Korpora umgehen können und eine nicht geleitete Herangehensweise schnell zu Demotivation aufgrund fehlender Ergebnisse führen kann. Gibt man den Suchausdruck nicht entsprechend der Programmiersprache des Korpus ein, führt die Suche zu keinem Ergebnis. Dabei reicht oft eine fehlende Leerstelle oder eine fehlende Klammer. Aus diesem Grund muss die Lehrkraft den Schülern erklären, worum es sich bei einem Korpus handelt, sie in die Funktionsweise der Korpora einführen und ihnen zumindest zu Beginn für Fragen zur Verfügung stehen.

Im Hinblick auf Sprachlehrer sieht Hunston zwei Haupteffekte, welche die Entwicklung von Korpora auf deren Berufsleben hat. Zum einen führen die Korpora zu einer neuen Beschreibung der jeweiligen Sprache, sodass sich auch der Inhalt des Sprachunterrichts durch sie verändert¹²⁹⁷, wie Sinclair es treffend formuliert: „The language looks rather different when you look at a lot of it at once.“¹²⁹⁸ Zum anderen können Korpora auch verwendet werden, um Materialien für den Fremdsprachenunterricht zu erstellen, und können somit die Basis für neue Ansätze der Lehrplanerstellung und der Methodologie legen.¹²⁹⁹

Wie am Vergleich der Verwendung des *subjunctivo* im Lehrbuch *¡Adelante! Nivel intermedio* und den Korpora CREA und CdE:Orig in Kapitel 4.2.5. gesehen werden konnte, können sich Untersuchungen in Korpora gewinnbringend auf die Auswahl von zu vermittelnden Unterrichtsinhalten auswirken. McEnery/Wilson weisen ebenso auf die falsche Darstellung von Aspekten des Sprachgebrauchs oder sprachlicher Variation in Textbüchern hin, welche ohne die Befragung von Korpora erstellt wurden. Daher betonen sie, dass bei der Entwicklung von Lehrbüchern Korpusstudien berücksichtigt werden sollten, um sicherzustellen, dass dem tatsächlichen Sprachgebrauch, also den unter Muttersprachlern üblichen Strukturen, Vorrang gewährt wird.¹³⁰⁰

Korpora können uns Auskunft darüber geben, wie eine Sprache beschaffen ist, und das Hauptargument für die Verwendung eines Korpus besteht darin, dass es verlässlicher ist als die Intuition von Muttersprachlern. Obwohl ein Muttersprachler Erfahrung mit weitaus mehr sprachlichen Daten hat als das größte Korpus, bleibt ein großer Teil dieser

¹²⁹⁷ Vgl. Hunston (2002): op. cit., S. 137.

¹²⁹⁸ Sinclair (1991): op. cit., S. 100.

¹²⁹⁹ Vgl. Hunston (2002): op. cit., S. 137.

¹³⁰⁰ Vgl. McEnery/ Wilson (1996): op. cit., S. 104.

Erfahrung der Introspektion unzugänglich. Muttersprachliche Lehrer sind beispielsweise häufig nicht in der Lage, zu sagen, warum in einem speziellen Kontext eine bestimmte Formulierung einer anderen vorzuziehen ist. Sie geben in diesem Fall meist an, dass es ‚sich einfach besser anhört‘, was den Lerner allerdings irritiert. Im Hinblick auf Kollokationen, Frequenz, Prosodie und Phraseologie ist die Intuition kein guter Ratgeber, weshalb Korpora besonders in diesen Bereichen vorzuziehen sind. Manche Kollokationen sind leicht intuitiv zu erfassen, beispielsweise *cometer un crimen*, andere sind jedoch weniger transparent, wie *error garrafal*, und werden von Lernern deshalb nicht verwendet, weil kein Lehrbuchautor oder Lehrer sich der Kollokation derart bewusst ist, dass diese in ein Lehrwerk oder in Unterrichtsmaterialien Eingang finden würde. In Bezug auf die Häufigkeit von Lexemen wurde bereits gesagt, dass es für Muttersprachler relativ schwer ist, Aussagen darüber zu treffen. Bis auf sehr allgemeine Fälle, wie zum Beispiel die Distribution von *tener* und *poseer*, bei denen jedem Sprecher klar ist, dass Letzteres weitaus seltener verwendet wird, kann die Intuition der Muttersprachler keinerlei Auskunft zur Frequenz geben, und es sollte daher ein Korpus befragt werden, um aufschlussreiche Ergebnisse zu erzielen. Zudem bringen bestimmte Ausdrücke laut Hunston bisweilen eine positive oder negative Konnotation mit sich. Wenn Sprecher damit konfrontiert werden, sind sie sich dessen entweder nicht bewusst und streiten die Existenz derselben teilweise sogar ab oder sie werden sich dessen schlagartig bewusst und fragen sich, warum sie nicht bereits früher darauf gekommen sind. In diesen Bereichen liefert ein Korpus ganz klar mehr Details als die Intuition eines Muttersprachlers. Dennoch will Hunston die Intuition nicht vollkommen ausschließen. Obwohl man sich nicht zu sehr auf sie verlassen dürfe, ist sie doch wichtig, um aus einer Menge an spezifischen Informationen aus einem Korpus Verallgemeinerungen zu erschließen.¹³⁰¹

Während Häufigkeit und Authentizität oft als zwei der bedeutendsten Vorteile des Gebrauchs von Korpora gesehen werden, sind sie auch ein Angriffspunkt für die Kritik von Wissenschaftlern auf dem Gebiet der Sprachpädagogik, so McEnery/Xiao.¹³⁰² Sie führen Cook an, welcher der Meinung ist, dass die Korpusdaten den Sprachunterricht verarmen lassen, dadurch, dass sie häufigen Merkmalen unangemessene Bedeutung

¹³⁰¹ Vgl. Hunston (2002): op. cit., S. 20-22.

¹³⁰² Vgl. McEnery/ Xiao (2010): op. cit., S. 373.

5. Schlussfolgerungen: Möglichkeiten und Grenzen der Verwendung von Korpora für den Fremdsprachenunterricht

zuschreiben auf Kosten seltener, aber effektiverer oder bedeutenderer Ausdrücke.¹³⁰³ In Bezug auf die Authentizität gibt es, wie bereits in Kapitel 3.2.1. zu sehen war, einige kritische Stimmen, wie beispielsweise Widdowson, nach dessen Meinung die Intention beim Lesen, also die Beziehung zwischen dem Text und dem Leser, eine große Rolle spielt und nicht unbedingt die Herkunft der Texte. Seiner Ansicht nach kann es sich bei den im Unterricht eingesetzten Texten zwar um reale Texte handeln, doch die Situation im Klassenzimmer ist eine andere als diejenige eines Muttersprachlers, der einen Zeitungsartikel aus Interesse liest. Die Verwendung des Texts ist somit nicht real.¹³⁰⁴

Osborne hingegen argumentiert, dass die Nutzung von Korpora auch die Chancen der Lerner erhöht, relativ seltenen Fällen des Sprachgebrauchs zu begegnen. Zudem können sie sehen, warum bestimmte Gebrauchsmuster atypisch sind, in welchen Kontexten sie vorkommen und wie sie sich in das Muster prototypischer Gebrauchsweisen einfügen.¹³⁰⁵

Auch Goethals sieht die Nutzung von Korpora für den Unterricht als vorteilhaft. Er prognostiziert für die Zukunft, dass die Einstufung nach der Häufigkeit ein Parameter für die Abfolge und Klassifizierung von Unterrichtsmaterialien sein wird, weil die Häufigkeit ein Maßstab für die Wahrscheinlichkeit der Nützlichkeit ist und hochfrequente Lexeme eine Art Kernvokabular darstellen, das sinnvoll ist im Hinblick auf die ansonsten zufällige Textauswahl eines Lehrers oder Lehrbuchautors.¹³⁰⁶

Obwohl allerdings Informationen zur Häufigkeit auf sehr einfache Art und Weise aus Korpora extrahiert werden können, hat noch kein Korpuslinguist behauptet, dass das Häufigste zugleich das Wichtigste ist. Im Gegenteil, so argumentiert Kennedy, ist es vielmehr so, dass die Häufigkeit nur eines von mehreren Kriterien sein sollte, das im Unterricht zu berücksichtigen ist, und dass die Fakten über die Sprache und den Sprachgebrauch, die aus Korpusanalysen ersichtlich werden, nie eine Bürde für die Pädagogik sein sollten. So ist es in manchen Bereichen sinnvoller, wenig frequenten Einheiten im Unterricht mehr Bedeutung zu schenken als sehr häufigen, weil über erstere

¹³⁰³ Vgl. Cook, Guy (1998): „The uses of reality: A reply to Ronald Carter“. In: *ELT Journal*, 52 (1), S. 61.

¹³⁰⁴ Vgl. Widdowson (1979): op. cit., S. 80-81.

¹³⁰⁵ Vgl. Osborne (2003): op. cit., S. 486.

¹³⁰⁶ Vgl. Goethals, Michaël (2003): „E.E.T.: the European English Teaching vocabulary-list“. In: Lewandowska-Tomaszczyk, B. (Hrsg.) (2003): *Practical Applications in language and computers*. Frankfurt: Peter Lang, S. 417-427.

bekannt ist, dass sie eine gewisse Schwierigkeit für Lerner darstellen.¹³⁰⁷ Es wäre jedoch unangemessen, zu sagen, dass Sprachlehrer, Lehrplandesigner und Materialersteller die vorhandenen Belege zur Häufigkeit ignorieren sollten, denn egal, warum die Gleichung ‚das Häufigste = das Wichtigste‘ im Bereich des Sprachlernprozesses falsch ist, so ist es doch schwer zu leugnen, dass aus Korpora entnommene Informationen zur Häufigkeit ein wichtiges empirisches Input für Sprachlernmaterialien darstellen.¹³⁰⁸

Als ein Nachteil der Verwendung von Korpora im Unterricht wird häufig die bereits erwähnte Komplexität der Eingabesprache angeführt. Nur nach einer gewissen Zeit intensiver Beschäftigung mit den Korpora ist es den Schülern möglich, auch komplexere Suchanfragen zu formulieren und die Ergebnisse richtig auszuwerten. Die Schüler werden im Verlauf der Suche bisweilen auch demotiviert aufgrund ausbleibender Ergebnisse. Dies geschieht meist, wenn die Suchanfragen entweder zu einfach oder zu komplex sind. Im ersten Fall kann die Suche unter Umständen sehr lange dauern, da viele Ergebnisse reproduziert werden müssen. Außerdem macht eine hohe Trefferzahl die Auswertung unübersichtlich. Im zweiten Fall kann das System häufig gar keine oder nur wenige Ergebnisse finden, was eine sinnvolle Auswertung unmöglich macht.¹³⁰⁹ Das Argument der Komplexität der Eingabesprache gilt jedoch hauptsächlich für ältere Korpora. In den neueren Korpora wie dem CORPES (siehe Kapitel 2.1.1.2.) oder dem CdE:Orig (siehe Kapitel 2.1.1.3.) kann anhand des vereinfachten und benutzerfreundlichen Interfaces sowohl mit einfachen Lexemen oder Phrasen als auch mit regulären Ausdrücken (siehe Kapitel 1.4.4.) gesucht werden. Eine Einführung durch die Lehrkraft sowie einige gemeinsame Suchabfragen sind zwar unabdingbar, doch die Schüler können nach einigen Sitzungen sehr wohl selbstständig im Korpus suchen.

Falls nicht genügend Ergebnisse im Korpus gefunden werden, so Bernardini, könne man nach alternativen Schreibweisen, nach Singular- und Pluralformen oder anderen Flexionsformen suchen. Des Weiteren können komplexe Ausdrücke in deren Einzelbestandteile zerlegt werden. Im gegenteiligen Fall kann die Suchanfrage durch die Angabe der Wortart oder des Kontexts weiter eingegrenzt werden. Zudem kann die Suche auf bestimmte Textsorten beschränkt werden.¹³¹⁰

¹³⁰⁷ Vgl. Kennedy (1998): op. cit., S. 290.

¹³⁰⁸ Vgl. Leech (1997): op. cit., S. 16.

¹³⁰⁹ Vgl. Bernardini (2000): op. cit., S. 230.

¹³¹⁰ Vgl. Bernardini (2000): op. cit., S. 232.

5. Schlussfolgerungen: Möglichkeiten und Grenzen der Verwendung von Korpora für den Fremdsprachenunterricht

Bernardini fand in ihrer Unterrichtseinheit zum Gebrauch großer Korpora im Englischunterricht heraus, dass Schüler häufig komplexe Suchanfragen formulieren konnten. Sie waren unter anderem in der Lage, Kollokationen zu ermitteln, konnten lexikalische Beobachtungen machen und fanden Erklärungen, Erweiterungen und Definitionen. Den Schülern fiel es leicht, Unterschiede und Übereinstimmungen bezüglich der Textsorte sowie der Position und Funktion von Ausdrücken im Text zu ermitteln. Auf der anderen Seite konnten sie selten Auskunft über die Repräsentativität von Beispielen geben, sie waren nur manchmal dazu in der Lage, lexikogrammatikalische Beobachtungen zu machen, sich alternative Suchstrategien auszudenken und genaue Zahlen anzugeben. Des Weiteren zogen sie nur selten andere Varianten eines Lexems heran oder reflektierten über semantische Beziehungen zwischen verschiedenen Lexemen.¹³¹¹ Auch aus eigenen Erfahrungen der Nutzung von Korpora mit Schülern kann gesagt werden, dass diese nach ein paar gemeinsamen Suchen durchaus in der Lage sind, Suchausdrücke zu formulieren und in einem Korpus Ergebnisse zu bestimmten Fragestellungen zu ermitteln. Generell fällt es den Schülern leicht, Kollokationen und Phraseologismen zu erkennen. Es gelingt ihnen ebenso, Präpositionen zu bestimmten Verben oder Substantiven herauszufinden oder Homonymen je nach der Bedeutung das jeweils richtige Genus zuzuordnen. Wortfamilien und Verbformen unterschiedlicher Tempora konnten ohne Probleme ausfindig gemacht werden, doch die Ermittlung der spanischen Pendanten des deutschen Verbs ‚werden‘ in Arbeitsblatt 11 (siehe Anhang 20) bereitete einigen Schülern Schwierigkeiten. Dieses Arbeitsblatt ist jedoch schließlich ebenso erfolgreich gelöst worden. Im Allgemeinen schätzten die Schüler besonders die vorbereiteten Wortschatzlisten wie in den Arbeitsblättern 26 bis 28, anhand derer sie sich vorab hinsichtlich des benötigten Vokabulars auf die Lektion vorbereiten konnten. Auf diese Listen gab es äußerst positive Rückmeldungen und insbesondere leistungsschwächere Schüler konnten erkennen, dass das notwendige Vokabular zum Verständnis der Texte der Lektion überschaubar ist. Zudem erleichterte die Verwaltung aller bislang gelesenen Texte in einem eigenen Korpus die Ermittlung unbekannten Vokabulars sowohl im Hinblick auf neue Texte, die im Unterricht gelesen werden sollten, als auch bezüglich der Erstellung von Klassenarbeiten, für welche die Angabe des unbekannten Wortschatzes besonders

¹³¹¹ Vgl. Bernardini (2000): op. cit., S. 231.

bedeutsam ist. Auch Beispielsätze für die Wiederholung von Grammatikeinheiten oder zur Veranschaulichung der Nutzung von Vokabular konnten mithilfe dieses Korpus auf relativ einfache Art und Weise aufgelistet werden.

Osborne gibt als einen weiteren Vorteil der Arbeit mit Korpora die Förderung der Neugier des Lernalers an. Die Wissbegierde der Schüler führt sie durch das Korpus und lässt sie nach interessanten Sachverhalten suchen, welche sie sich selbst erklären möchten. Sie suchen beispielsweise nach Lösungen für Schwierigkeiten, welche sie selbst während der Produktion eigener Texte hatten oder nach sprachlichen Besonderheiten. Die Schüler verbringen seiner Meinung nach eine beachtliche Zeit mit der Suche im Korpus und der Analyse von Daten. Des Weiteren äußern sie sowohl ihre Frustration darüber, dass ihnen das Korpus nicht sofort die gewünschten Antworten präsentiert, als auch ihre Zufriedenheit darüber, dass sie mit realen Sprachdaten arbeiten.¹³¹² De Jacobi berichtet ebenso, dass ihre korpusbasierte Unterrichtseinheit zwar für ein paar Schüler kompliziert, konfus und ermüdend war, für einige jedoch war es eine alternative Art des Lernens, die es ihnen erlaubte, in direkten Kontakt mit realen Sprachdaten zu treten, zur Reflexion anregte und das Lernen erleichterte.¹³¹³ Auch diese Beobachtungen können bestätigt werden. Zunächst haben einige Schüler Probleme mit der Vorgehensweise bei Suchen im Korpus und zeigen durchaus ihre Unzufriedenheit über erfolglose Suchanfragen, doch nach einigen erfolgreichen Ergebnissen werden sie immer interessierter und stellen sogar eigene Nachforschungen an, welche über die Aufgaben auf dem Arbeitsblatt hinaus gehen. Besonders interessierte Schüler sind vor allem davon begeistert, dass sie in realen Sprachdaten forschen und anhand dieser herausfinden können, wie Muttersprachler gewisse Sachverhalte formulieren.

Osborne schreibt in seinem Aufsatz über die Integration von Korpora in einen Fremdsprachenlehrplan allerdings, dass es für Lerner schwierig ist, mit Korpora gesprochener Sprache zu arbeiten, da sie die Texte aufgrund der vielen Wiederholungen, Verzögerungen und Reformulierungen nicht entsprechend analysieren und verstehen können. Durch die ständigen Unterbrechungen werden sie verwirrt, erkennen vorhandene Nebensätze, wie beispielsweise Relativsätze, nicht und können dadurch manche Textausschnitte nicht interpretieren.¹³¹⁴ Für das eigene Forschen von Schülern im Korpus

¹³¹² Vgl. Osborne (2003): op. cit., S. 490-491.

¹³¹³ Vgl. De Jacobi (2002): op. cit., S. 42.

¹³¹⁴ Vgl. Osborne (2003): op. cit., S. 479-481.

5. Schlussfolgerungen: Möglichkeiten und Grenzen der Verwendung von Korpora für den Fremdsprachenunterricht

sind sicherlich Korpora der geschriebenen Sprache vorzuziehen. Dennoch kann die Lehrkraft Daten aus Korpora gesprochener Sprache nutzen, beispielsweise für Hörverstehensaufgaben oder für Arbeitsblätter, doch in letzterem Fall sollte sie überflüssige Wiederholungen löschen, um die Schüler nicht unnötig zu verwirren. Zudem sind die Transkripte in einigen der Korpora gesprochener Sprache oder auch der multimodalen Korpora so gestaltet, dass sie sich wie ein geschriebener Text lesen, da Wiederholungen und Verzögerungen gelöscht wurden.

Braun/Chambers sehen die Nutzung von Korpora für den Unterricht durchweg positiv, wobei bei der Erstellung von Arbeitsblättern mit Konkordanzen eine gute Planung durch die Lehrperson vorausgesetzt wird. Bei weniger fortgeschrittenen Lernern sollte viel Wert auf Verständlichkeit und eine klare Fokussierung der Konkordanzen auf das jeweilige Thema geachtet werden.¹³¹⁵ Sollte die direkte Nutzung der Korpora am Computer jedoch zu schwierig für die Lerner sein, so kann die Lehrkraft Arbeitsblätter mit Korpusmaterial erstellen und den Schülern als Übung vorlegen, wie beispielsweise bei der Fehlerkorrektur in Kapitel 4.2.6. oder wie in Kapitel 4.2.4. mit Konkordanzen aus dem Korpus *Sacodeyl*. Vor allem im Anfangsstadium und mit jüngeren Schülern ist diese Vorgehensweise sehr vorteilhaft.

Generell gilt es zu bedenken, dass die Arbeit mit Korpora nur eine im Unterricht verwendete Methode sein sollte, denn wie Haß es formuliert:

„Mit Blick auf die Effizienz unterschiedlicher methodischer Verfahren weist die neuere empirische Unterrichtsforschung darauf hin, dass keine Methode *per se* zu besonderen Lernergebnissen führt, sondern dass die planmäßige, zielgerichtete, situations- und adressatenadäquate Auswahl der Methoden das entscheidende Qualitätskriterium für guten (Fremdsprachen-)Unterricht darstellt. Dabei ist ein Optimum an Methodenvariation (keine methodische Monokultur und kein hektischer Methodenmaximalismus) nachgewiesenermaßen am förderlichsten.“¹³¹⁶

Die in dieser Arbeit dargestellten Möglichkeiten der Nutzung von Korpora können weder in ihrer Fülle in Lehrbücher integriert werden noch sollen sie das Lehrbuch oder traditionelle Unterrichtsmethoden ersetzen, sie sollen lediglich Vorschläge darstellen, wie linguistische Korpora für den Unterricht genutzt werden und diesen durch eine neue Herangehensweise auflockern können. Es soll damit veranschaulicht werden, welche Alternativen die Korpora für den Fremdsprachenunterricht bieten und es soll zur weiteren

¹³¹⁵ Vgl. Braun/ Chambers (2009): op. cit., S. 334.

¹³¹⁶ Haß, Frank (2013): „Methoden im Fremdsprachenunterricht“. In: Hallet, W./ Königs, F. (Hrsg.) (2013): *Handbuch Fremdsprachendidaktik*. Seelze-Velber: Kallmeyer [u.a.], S. 155.

Beschäftigung mit den Korpora der spanischen Sprache sowohl durch die Lehrkraft als auch durch Schüler angeregt werden.

Alonso Pérez-Ávila führt zudem an, dass weder Lehrer noch Schüler noch Schulen derzeit dazu bereit sind, die Struktur des traditionellen Unterrichts hinter sich zu lassen. In den vergangenen Jahren hat es große Veränderungen gegeben und die Schulen passen sich nach und nach daran an. Ihrer Meinung nach werden die Schüler bei der Annäherung an ihre aktive Rolle im Lernprozess zunächst ein wenig verwirrt und verloren sein, da die bisherige Rolle des Lehrers, der alle Zweifel auslöscht und in jedem Moment angibt, was zu tun ist, verschwindet. Sie werden verunsichert sein, wenn sie die Lehrkraft darum bittet, die bisher fest geltenden Klassenregeln hinter sich zu lassen, und sie werden einige Zeit brauchen, um sich daran zu gewöhnen, dass sie nun die wahren Protagonisten ihres eigenen Lernprozesses sind und anfangen müssen, als aktive Teilnehmer an diesem Prozess mitzuwirken.¹³¹⁷

Braun/Chambers sind allerdings der Meinung, die mangelnden Korpuskenntnisse der Lehrkraft seien ein wichtiger Faktor, welcher derzeit gegen den Einsatz von Korpora im Unterricht spricht. Sie sehen daher die Hochschulausbildung sowie Lehrerfortbildungen als Angriffspunkte für eine Veränderung der aktuellen Situation.¹³¹⁸ Auch für McEnery/Xiao stellt die mangelnde Ausbildung der Lehrkräfte im Bereich der Arbeit mit Korpora ein Problem für die Nutzung derselben im Unterricht dar.¹³¹⁹ Leech geht sogar so weit, dass in Sprachkursen auf universitärem Niveau neben den standardisierten Aufgabenstellungen wie dem Verfassen kritischer und argumentativer Aufsätze auch Aufgaben gestellt werden sollten, bei denen die Lerner nach eigenem Ermessen (nicht nur aber zumindest teilweise unter Zuhilfenahme von Korpora) reale Sprachdaten organisieren und untersuchen sollen, sodass zwischen diesen Aufgabentypen ein Gleichgewicht herrscht.¹³²⁰ Wenn korpusbasierte Aufgaben Eingang in den Unterricht an weiterführenden Schulen und nicht nur an Universitäten finden sollen, sind diese noch mehr von der Vermittlung durch die Lehrkraft abhängig. Aus eben diesem

¹³¹⁷ Vgl. Alonso Pérez-Ávila (2006): op. cit. (06.06.2017), S. 76.

¹³¹⁸ Vgl. Braun/ Chambers (2009): op. cit., S. 336.

¹³¹⁹ Vgl. McEnery/ Xiao (2010): op. cit., S. 374-375.

¹³²⁰ Vgl. Leech (1997): op. cit., S. 10.

5. Schlussfolgerungen: Möglichkeiten und Grenzen der Verwendung von Korpora für den Fremdsprachenunterricht

Grund sind die in Kapitel 3.3.1. vorgestellten direkten Verwendungsweisen der Korpora im Unterricht zumindest zum Teil derzeit noch nicht möglich.

Im Hinblick auf die angesprochenen Voraussetzungen an den Schulen gilt es laut Alonso Pérez-Ávila und Siepmann auch, die technische Ausstattung derselben zu beachten.¹³²¹ Die Anzahl von Computern mit Internetzugang ist an jeder Schule unterschiedlich. Um autonomes Lernen und selbstständiges Forschen halbwegs gewährleisten zu können, sollte bei der Arbeit mit Korpora mindestens ein Computer für je zwei Schüler bereitstehen. Bei der Verfügbarkeit von Computern handelt es sich um ein generelles Problem, das es für die Mehrheit der Schulen zu bedenken gilt. Wären an öffentlichen Schulen Computer mit Internetzugang in allen Klassenzimmern verfügbar, könnten Korpora schnell, einfach und auch spontan im Unterricht genutzt werden, wenn während des Unterrichts Fragen zu Wortschatz oder Grammatik aufkommen.

Siepmann ist außerdem der Meinung, dass neben der schlechten technischen Ausstattung der Schulen vor allem der Mangel an Korpora mit pädagogischen Zielsetzungen gegen die Verwendung von Korpora im Unterricht spricht. Bereits bestehende Korpora wie das *British National Corpus* enthielten reales Sprachmaterial, das insbesondere im Anfangsunterricht Verständnisprobleme bereitet, weshalb er die Erstellung von Korpora für das schulische Lernen fordert, welche relativ einfaches, aber dennoch reales Material bieten. Die Texte könnten beispielsweise aus Kinder- und Jugendzeitungen beziehungsweise –büchern entnommen werden.¹³²²

Braun/Chambers kritisieren ebenso den Mangel an geeigneten Korpora in vielen Sprachen. Ein im Unterricht nutzbares Korpus sollte einfach, kostenlos zugänglich sowie sprachdidaktisch relevant sein, doch diese Art von Korpora sind derzeit noch wenig verfügbar oder nicht leicht zu finden. Außerdem sehen sie weitere empirische Studien zum Einsatz von Korpora im Sekundarschulbereich als unabdingbar, um den Lehrkräften Anregungen zu geben.¹³²³

Auch McEnery/Xiao sehen neben der mangelnden Ausbildung der Lehrkräfte die derzeit verfügbaren Korpora als Hindernis für die Nutzung von Korpora im Unterricht. Die bislang zugänglichen Korpora sind in der Mehrheit zu sprachwissenschaftlichen Zwecken erstellt worden und daher für den Gebrauch im Unterricht nur teilweise

¹³²¹ Vgl. Alonso Pérez-Ávila (2006): op. cit. (06.06.2017), S. 77. Und: Siepmann (2009): op. cit., S. 328.

¹³²² Vgl. Siepmann (2009): op. cit., S. 328.

¹³²³ Vgl. Braun/ Chambers (2009): op. cit., S. 336.

geeignet. Somit müssten zunächst pädagogisch motivierte Korpora zusammengestellt werden, sowohl in Bezug auf das Design als auch hinsichtlich des Inhalts, um pädagogische Bedürfnisse und Anforderungen des Lehrplans zu befriedigen, sodass korpusbasierte Lernaktivitäten ein integraler Teil des allgemeinen Sprachlehrplans werden können anstatt lediglich eine zusätzliche Möglichkeit zu sein. Dabei muss außerdem dem Alter der Lerner, ihren Erfahrungen und ihrem sprachlichen Niveau sowie der Integration in den allgemeinen Lehrplan Rechnung getragen werden. McEnery/Xiao sind der Meinung, dass durch die Behebung dieser beiden Mängel, der bislang unzureichenden Lehrerausbildung im Bereich der Korpora sowie der mangelnden Verfügbarkeit pädagogisch motivierter Korpora, die Art und Weise, wie wir an die Spracherziehung herangehen, grundlegend verändert werden wird, sowohl im Hinblick darauf, was unterrichtet wird, als auch darauf, wie unterrichtet wird.¹³²⁴

Im englischsprachigen Raum ist die Nutzung von Korpora generell schon weit fortgeschritten, wie vor allem im dritten Kapitel zu sehen war, doch auch hier fehlt es an pädagogisch motivierten Korpora, wie McEnery/Xiao¹³²⁵ sie fordern. Es wäre für die Forschung zur spanischen Sprache sowie für den Spanischunterricht damit eine äußerst positive Entwicklung, wenn auch mehr spanische Korpora erstellt und diese didaktisch genutzt würden.

McEnery/Wilson sehen beispielsweise Vorteile bei der Nutzung von Parallelkorpora im Bereich der Übersetzung, da die Schüler darin parallel Stilbeispiele und Phraseologismen in mehr als einer Sprache betrachten oder ihre eigene Übersetzung mit einer professionellen vergleichen können.¹³²⁶ Derzeit gibt es allerdings kein spanisch-deutsches Parallelkorpus, wie in Kapitel 2.3. zu sehen war. Ein solches Korpus würde nicht nur die Hispanistik allgemein voranbringen, sondern auch einige Vorteile für die Übersetzung im Spanischunterricht in der Schule bieten. Zu schulischen Zwecken wäre hierbei jedoch eine Art pädagogisches Parallelkorpus wünschenswert, das Texte auf angemessenem Niveau enthält.

Insbesondere ein großes pädagogisches Korpus ähnlich dem *Sacodeyl*, welches Texte unter Angabe des Themas, des sprachlichen Niveaus und darin vorkommender Grammatikeinheiten enthält, wäre für die Zukunft wünschenswert. Es wäre ebenso

¹³²⁴ Vgl. McEnery/ Xiao (2010): op. cit., S. 374-375.

¹³²⁵ Vgl. McEnery/ Xiao (2010): op. cit., S. 374-375.

¹³²⁶ Vgl. McEnery/ Wilson (1996): op. cit., S. 106.

5. Schlussfolgerungen: Möglichkeiten und Grenzen der Verwendung von Korpora für den Fremdsprachenunterricht

förderlich, wenn es sich nicht nur um Interviews mit spanischen Jugendlichen, sondern auch um Gespräche handeln würde. Anhand eines derartigen pädagogischen Korpus könnten speziell auf die Lerngruppe abgestimmte Texte ausgewählt und im Unterricht für die Ermittlung landeskundlicher Informationen oder für ein Hörverstehen genutzt werden. Es wäre außerdem von Vorteil, wenn es ein Korpus der geschriebenen Sprache gäbe, welches Texte unterschiedlicher Kompetenzniveaus zu den für den Unterricht laut GER (siehe Kapitel 3.1.2.) und Lehrplan (siehe Kapitel 3.1.3.) relevanten Themen enthielte. Für den schulischen Fremdsprachenunterricht wären beispielsweise Texte wie Briefe, E-Mails oder Postkarten von Nutzen. Ähnlich wie bei den Aufgaben zum Hörverstehen (siehe Kapitel 4.4.1.) könnten anhand eines solchen Korpus auch Aufgaben zum Leseverstehen erstellt werden. Texte aus Referenzkorpora könnten für derartige Aufgaben ebenso verwendet werden, doch in den gängigen Referenzkorpora sind die vollständigen Texte nicht zugänglich, wie in Kapitel 2.1.1. zu sehen war. Zudem sind Dateien aus diesen Korpora sehr schwer verständlich und können daher erst ab dem 4. Lernjahr verwendet werden. Ein Vorbild für die Erstellung eines solchen pädagogischen Korpus wäre das in Kapitel 3.3.2.2.1. beschriebene CRFC, welches einen beachtlichen Teil gesprochener Texte aus Transkripten von Fernsehsendungen umfasst. Ein für schulische Zwecke gut nutzbares Korpus sollte zu einem großen Anteil Daten gesprochener Sprache in sich vereinen sowie Texte auf einem angemessenen sprachlichen Niveau enthalten, welche sich mit den kommunikativen Lernzielen des Spanischunterrichts befassen, wie von der Redaktion Spanisch/Italienisch/Russisch des Cornelsen Verlags gefordert (siehe Kapitel 3.3.2.2.1.).

So gilt es mit Blick auf die Zukunft zu hoffen, dass mehr spanische, insbesondere pädagogisch motivierte und didaktisch nutzbare Korpora erstellt werden, sich die technische Ausstattung der Schulen verbessert und die Ausbildung der Lehrkräfte aufgrund der zunehmenden Digitalisierung auch im Hinblick auf die Nutzung von Korpora für und im Unterricht ausgebaut wird. Dennoch konnte in den vorherigen Kapiteln gezeigt werden, dass auch die bisher verfügbaren Korpora bereits auf unterschiedlichste Art und Weise im und für den Spanischunterricht genutzt werden können, sei es als Stütze für die Lehrkraft bei der Korrektur von Schülerarbeiten oder zum Abgleich bereits bestehender Unterrichtsmaterialien, als Quelle für reale Texte und Beispielsätze für Klassenarbeiten und Arbeitsblätter oder aber als auf vielfältige Weise

durchsuchbare Textsammlung für Schüler sowohl zur Ermittlung von Kollokationen, Synonymen oder Wortfamilien als auch zur Verbesserung von Textproduktionen. Für besonders versierte Lehrkräfte kann die Erstellung eines eigenen Korpus zudem Vorteile für die Vorbereitung von Unterrichtsmaterialien und den Abgleich mit neuen Texten bringen. Korpora sind damit ein nicht zu unterschätzendes Instrument für den Fremdsprachenunterricht, dem mehr Beachtung geschenkt werden sollte.

Mit der vorliegenden Arbeit ist nun der Grundstein für die Nutzung spanischer Korpora im Sprachunterricht an deutschen Schulen gelegt, denn sie stellt, wie zu sehen war, eine Art Leitfaden dar, der von interessierten Lehrkräften herangezogen werden kann, sofern sie sich näher mit linguistischen Korpora beschäftigen möchten. Außerdem soll die Dissertation speziell für den Einsatz in der Schule aufbereitet und inklusive der Arbeitsblätter in einem Verlag veröffentlicht werden, um es allen Lehrkräften zu ermöglichen, sich ein Bild von der Korpuslinguistik an sich, vor allem aber vom Nutzen der Korpora für den eigenen Unterricht zu machen. Die angesprochene Publikation soll zudem eine detaillierte Anleitung für die eigene Erarbeitung eines Korpus, wie sie im Exkurs zum 4. Kapitel kurz beschrieben wurde, enthalten, um den Lehrkräften Schritt für Schritt zu zeigen, wie ein eigenes Korpus erstellt werden kann, und sie auf eventuelle Problemfälle hinzuweisen. Es bleibt somit zu hoffen, dass einige Lehrer dazu motiviert werden, Korpora im und für ihren Unterricht zu verwenden, um auf diese Weise reale Sprachmodelle ins Klassenzimmer zu bringen oder die Schüler sogar selbst darin forschen zu lassen.

6. Bibliographie

6.1. Nachschlagewerke

Amann, Klaus/ Marín Barrera, Sara/ Schleyer, Jochen/ Vicente Álvarez, Araceli/
Wlasak-Feik, Christine (2007): *Encuentros 2. Método de español*. Berlin: Cornelsen, S.
54.

Arellano, Jesús (2006): *Langenscheidt sm Diccionario de Español para Extranjeros*.
München [u.a.]: Langenscheidt.

Arriagada Espinoza, Melanie (2012): *¡Adelante! Nivel avanzado. Schülerbuch*. Stuttgart:
Klett.

Barquero, Antonio/ Collado Revestido, Cristina/ Corpas, Jaime/ Dienhoff, Oliver/
Domínguez, Yolanda/ Kuhlmann, Erika/ Navarro, Javier/ Reiter, Claudia (2010):
¡Adelante! Nivel elemental. Schülerbuch. Stuttgart: Klett.

Barquero, Antonio/ Bizama, Liliana/ Corpas, Jaime/ Crovetto-Bizama, Pedro/ Díaz
Gutiérrez Eva/ Jeske, Claire-Marie/ Jiménez Romera, Alicia/ Navarro, Javier/
Pardellas Velay, Rosamna (2011): *¡Adelante! Nivel intermedio. Schülerbuch*.
Stuttgart: Klett.

Bosque, Ignacio (2004): *REDES. Diccionario combinatorio del español contemporáneo*.
Madrid: Ediciones SM.

Bosque, Ignacio (2006): *Diccionario combinatorio práctico del español contemporáneo*.
Las palabras en su contexto. Madrid: Ediciones SM.

Bosque, Ignacio/ Demonte, Violeta (Hrsg.) (1999): *Gramática descriptiva de la lengua
española*. 3 vols. Madrid: Espasa.

Bußmann, Hadumod (2008): *Lexikon der Sprachwissenschaft*. Stuttgart: Kröner.

Butt, John/ Benjamin, Carmen (1988): *A New Reference Grammar of Modern Spanish*.
London: Arnold.

Davies, Mark (2006): *A Frequency Dictionary of Spanish. Core Vocabulary for Learners*.
New York: Routledge.

De Bruyne, Jacques (2002): *Spanische Grammatik*. Tübingen: Max Niemeyer Verlag, 2.
Auflage.

Deuter, Margaret (2002): *Oxford Collocations Dictionary for Students of English*.
Oxford: Oxford University Press.

- Dorn, Rudolf/ Navarro González, Javier/ Strathmann, Jochen (2012): *¡Gramática! de la lengua española*. Stuttgart: Klett.
- Dudenredaktion (ed.2014): *Duden – Das Synonymwörterbuch. Ein Wörterbuch sinnverwandter Wörter*. Berlin: Dudenverlag.
- Görrissen, Margarita (2005): *Praxis-Grammatik Spanisch*. Stuttgart: Klett.
- Gutiérrez Cuadrado, Juan (2006): *Diccionario Salamanca de la lengua española*. Madrid: Santillana [u.a.].
- Huber, Margit/ Mancini, Lucia/ Simó Perales Zabka, Susana/ Vega Ordóñez, Silvia (2011): *¡Adelante! Nivel intermedio. Lehrerbuch mit Lehrersoftware und Video-DVD*. Klett: Stuttgart.
- Hueber, Margit/ Vega Ordoñez, Silvia (2010): *¡Adelante! Nivel elemental. Lehrerbuch mit Lehrersoftware und Video-DVD*. Klett: Stuttgart.
- Köck, Peter (2008): *Wörterbuch für Erziehung und Unterricht*. Augsburg: Brigg Pädagogik Verlag.
- Mindt, Dieter (1995): *An empirical grammar of the English verb: modal verbs*. Berlin: Cornelsen.
- Mindt, Dieter (2000): *An empirical grammar of the English verb system*. Berlin: Cornelsen.
- Rathsam, Kathrin/ Schleyer, Jochen (2014): *Spanische Grammatik für die Mittel- und Oberstufe*. Berlin: Cornelsen.
- Reumuth, Wolfgang/ Winkelmann, Otto (2011): *Praktische Grammatik der spanischen Sprache*. Wilhelmsfeld: Gottfried Egert Verlag.
- Sebastián Gallés, Núria/ Cuetos Vega, Fernando/ Carreiras Valiña, Manuel Francisco/ Martí Antonin, Maria Antònia (2000): *LEXESP. Léxico informatizado del español*. Barcelona: Edicions de la Universitat de Barcelona.
- Schmidt, Ulrike (2005): *Wortschatz zum Verstehen: Französisch. Nützliches Wörterbuch für fortgeschrittene Lerner*. Aachen: Shaker.
- Süß, Kurt/ Pérez, Petronilo/ Ruipérez, Germán (2003): *Lerngrammatik Spanisch*. Braunschweig: Diesterweg.
- Worsch, Wolfgang (2012): *Cornelsen Klausurwörterbuch Spanisch*. Berlin: Cornelsen.

6.2. Sekundärliteratur

- Alcántara Plá, Manuel (2007): *Introducción al análisis de estructuras lingüísticas en corpus. Aproximación semántica*. Madrid: UAM.
- Alderson, Charles (1996): „Do corpora have a role in language assessment?“. In: Thomas, J./ Short, M. (Hrsg.) (1996): *Using corpora for language research*. London: Longman, S. 248–259.
- Allan, Quentin Grant (1999): „Enhancing the language awareness of Hong Kong teachers through corpus data“. In: *Journal of Technology and Teacher Education*, 7 (1), S. S. 57–74.
- Alonso-Ramos, Margarita (Hrsg.) (2016): *Spanish Learner Corpus Research: Current trends and future perspectives*. Amsterdam [u.a.]: John Benjamins.
- Alonso-Ramos, Margarita (2016): „Spanish learner corpus research: Achievements and challenges“. In: Alonso-Ramos, M. (Hrsg.) (2016): *Spanish Learner Corpus Research: Current trends and future perspectives*. Amsterdam [u.a.]: John Benjamins, S. 3-31.
- Alvarado Ortega, María Belén (2010): *Las fórmulas rutinarias del español: teoría y aplicaciones*. Frankfurt am Main: Lang.
- Ammermann, Anne (2013): „English-Spanish code-switching in the Facebook communication of Arkansas authors“. In: Hennemann, A./ Schlaak, C. (Hrsg.) (2013): *Korpuslinguistische Untersuchungen: Analysen einzelsprachlicher Phänomene*. Berlin: Frank & Timme, S. 187-198.
- Aston, Guy (2000): „Corpora and language teaching“. In: Burnard, L./ McEnery, T. (Hrsg.) (2000): *Rethinking language pedagogy from a corpus perspective: papers from the Third International Conference on Teaching and Language Corpora*. Frankfurt am Main: Peter Lang, S. 7-17.
- Aston, Guy/ Bernardini, Silvia/ Stewart, Dominic (Hrsg.) (2004): *Corpora and Language Learners*. Amsterdam [u.a.]: Benjamins.
- Bahns, Jens (1997): *Kollokationen und Wortschatzlernen im Englischunterricht*. Tübingen: Gunter Narr.
- Barker, Fiona/ Salamoura, Angeliki/ Saville, Nick (2015): „Learner Corpora and Language Testing“. In: Granger, S./ Gilquin, G./ Meunier, F. (Hrsg.) (2015): *The Cambridge Handbook of Learner Corpus Research*. Cambridge: Cambridge University Press, S. 511-534.

- Barkowski, Hans (2007): „Zweitsprachunterricht“. In: Bausch, K.-R./ Christ, H./ Krumm, H.-J. (Hrsg.) (2007): *Handbuch Fremdsprachenunterricht*. Tübingen/ Basel: A. Francke, S. 157.
- Baroni, Marco/ Bernardini, Silvia (Hrsg.) (2006): *WaCky! Working papers on the web as corpus*. Bologna: Gedit.
- Becker, Martin (2013): *Einführung in die spanische Sprachwissenschaft*. Stuttgart: Metzler.
- Benson, Morton (1985): „Collocation and idioms“. In: Ilson, R. (Hrsg.) (1985): *Dictionaries, lexicography and language learning*. Oxford: British Council/ Pergamon, S. 61-68.
- Bernardini, Silvia (2000): „Systematising serendipity: Proposals for concordancing large corpora with language learners“. In: Burnard, L./ McEnery, T. (Hrsg.) (2000): *Rethinking language pedagogy from a corpus perspective: papers from the Third International Conference on Teaching and Language Corpora*. Frankfurt am Main: Peter Lang, S. 225-234.
- Bernardini, Silvia (2004): „Corpora in the classroom. An overview and some reflections on future developments“. In: Sinclair, J. McH. (2004): *How to use corpora in language teaching*. Amsterdam/ Philadelphia: Benjamins, S. 15-36.
- Biber, Douglas (1994): „Representativeness in Corpus Design“. In: Zampolli, A./ Calzolari, N./ Palmer, M. (1994): *Current issues in computational linguistics. In honour of Don Walker*. Pisa: Giardini, S. 377-407.
- Biber, Douglas (2009): „A corpus-driven approach to formulaic language in English: multi-word patterns in speech and writing“. In: *International Journal of Corpus Linguistics* 14 (3), S. 275-311.
- Bickmann, Hans-Jürgen (1978): *Synonymie und Sprachverwendung: Verfahren zur Ermittlung von Synonymenklassen als kontextbeschränkten Äquivalenzklassen*. Tübingen: Niemeyer.
- Blanco Canales, Ana (2016): „Factors that can have an impact on the processes of perceiving Spanish/L2“. In: Alonso-Ramos, M. (Hrsg.) (2016): *Spanish Learner Corpus Research: Current trends and future perspectives*. Amsterdam [u.a.]: John Benjamins, S. 199-234.
- Blank, Andreas (2001): *Einführung in die lexikalische Semantik für Romanisten*. Tübingen: Niemeyer.
- Bludau, Michael (1996): „What is ‘authentic’?“. In: *Zielsprache Englisch*, 1996, 26 (2), S. 11-13.

6. Bibliographie

- Böhm, Verónica (2013): „Evidentielle Markierung von fremden Äußerungen. Eine Analyse des spanischen Imperfekts in journalistischen Texten“. In: Hennemann, A./ Schlaak, C. (Hrsg.) (2013): *Korpuslinguistische Untersuchungen: Analysen einzelsprachlicher Phänomene*. Berlin: Frank & Timme, S. 119-130.
- Borja, Anabel (2008): „Corpora for Translators in Spain. The CDJ-GITRAD Corpus and the GENTT Project“. In: Andermann, G./ Rogers, M. (Hrsg.) (2008): *Incorporating Corpora. The linguist and the translator*. Clevedon [u.a.]: Multilingual Matters, S. 243-265.
- Braun, Sabine/ Chambers, Angela (2009): „Elektronische Korpora als Ressource für den Fremdsprachenunterricht“. In: Jung, U. O. H. (Hrsg.) (2009): *Praktische Handreichung für Fremdsprachenlehrer*. Frankfurt am Main [u.a.]: Lang, S. 330-337.
- Breen, Michael P. (1985): „Authenticity in the Language Classroom“. In: *Applied Linguistics*, Vol. 6, No. 1, S. 60-70.
- Buckingham, Louisa (2009): *Las construcciones con verbo soporte en un corpus de especialidad*. Frankfurt am Main: Lang.
- Burnard, Lou/ McEnery, Tony (Hrsg.) (2000): *Rethinking language pedagogy from a corpus perspective: papers from the Third International Conference on Teaching and Language Corpora*. Frankfurt am Main: Lang.
- Burr, Elisabeth (2004): „Das Korpus romanischer Zeitungssprachen in Forschung und Lehre“. In: Dahmen, W./ Holtus, G./ Kramer, J./ Metzeltin, M./ Schweickard, W./ Winkelmann, O. (Hrsg.) (2004): *Romanistik und neue Medien*. Romanistisches Kolloquium XVI (=TBL 455). Tübingen: Narr, S. 133-162.
- Busse, Dietrich (2006): „Sprachnorm, Sprachvariation, Sprachwandel. Überlegungen zu einigen Problemen der sprachwissenschaftlichen Beschreibung des Deutschen im Verhältnis zu seinen Erscheinungsformen“. In: *Deutsche Sprache*, 34. Jg., Heft 4/2006, S. 314-333.
- Bustos Plaza, Alberto (2012): „Kollokationen und Funktionsverbgefüge“. In: Born, J./ Folger, R./ Laferl, C. F./ Pöll, B. (Hrsg.) (2012): *Handbuch Spanisch*. Berlin: Schmidt, S. 251-257.
- Cabrerizo Ruiz, María Aránzazu (2000): *Sueña 2*. Madrid: Anaya.
- Callies, Marcus/ Götz, Sandra (Hrsg.) (2015): *Learner Corpora in Language Testing and Assessment*. Amsterdam: John Benjamins.
- Campillos Llanos, Leonardo (2016): „POS-tagging a Spanish oral learner corpus: Criteria, procedure, and a sample analysis“. In: Alonso-Ramos, M. (Hrsg.) (2016): *Spanish Learner Corpus Research: Current trends and future perspectives*. Amsterdam [u.a.]: John Benjamins, S. 89-116.

- Carter, Ronald/ McCarthy, Michael (1995): „Grammar and the spoken language“. In: *Applied Linguistics*, 16 (2), S. 141–158.
- Cheng, Winnie/ Warren, Marten (2000): „The HongKong Corpus of Spoken English: language learning through language description“. In: Burnard, L./ McEnery, T. (Hrsg.) (2000): *Rethinking language pedagogy from a corpus perspective: papers from the Third International Conference on Teaching and Language Corpora*. Frankfurt am Main: Lang, S. 133-144.
- Ciapuscio, Guiomar (2007): „Epistemic modality and academic orality: Pilot study for COTECA (Corpus Textual del Español Científico de la Argentina)“. In: Parodi, G. (Hrsg.) (2007): *Working With Spanish Corpora*. London: Continuum, S. 90-105.
- Ciesielska-Ciupek, Maria (2003): „Teaching with the internet and corpus materials. Preparation of the ELT materials using the Internet and corpus resources“. In: Lewandowska-Tomaszczyk, B. (Hrsg.) (2003): *PALC 2001: Practical Applications in Language Corpora*. Frankfurt am Main: Lang, S. 521-531.
- Clear, Jeremy (2000): „Do you believe in Grammar?“. In: Burnard, L./ McEnery, T. (Hrsg.) (2000): *Rethinking language pedagogy from a corpus perspective: papers from the Third International Conference on Teaching and Language Corpora*. Frankfurt am Main: Peter Lang, S. 19-30.
- Cobb, Tom (1997): „Is there any measurable learning from hands on concordancing?“. In: *System*, 25(3), S. 301–315.
- Cobb, Tom/ Horst, Marlise (2001): „Reading academic English: carrying learners across the lexical threshold“. In: Flowerdew, J./ Peacock, M. (Hrsg.) (2001): *Research Perspectives on English for Academic Purposes*. Cambridge: Cambridge University Press. S. 315-329.
- Collins, Heloisa (2000): „Materials design and language corpora: a report in the context of Distance Education“. In: Burnard, L./ McEnery, T. (Hrsg.) (2000): *Rethinking language pedagogy from a corpus perspective: papers from the Third International Conference on Teaching and Language Corpora*. Frankfurt am Main: Peter Lang, S. 51-63.
- Coniam, David (1997): „A preliminary inquiry into using corpus word frequency data in the automatic generation of English language cloze tests“. In: *CALICO Journal* 16 (2–4), S. 15–33.
- Conolly, Declan B. (1996): „Are there degrees of authenticity?“. In: *Zielsprache Englisch*, 1996, 26 (2), S. 15-18.
- Conrad, Susan (1999): „The importance of corpus-based research for language teachers“. In: *System*, 27, S. 1–18.

6. Bibliographie

- Cook, Guy (1998): „The uses of reality: A reply to Ronald Carter“. In: *ELT Journal*, 52 (1), S. 57-64.
- Corpas Pastor, Gloria (1992): „Tratamiento de las colocaciones del tipo A+S/S+A en diccionarios bilingües y monolingües (español-inglés)“. In: Alvar Ezquerro, M. (Hrsg.) (1992): *EURALEX '90 Proceedings. Actas del IV Congreso Internacional*. Barcelona: Bibliograf, S. 331-340.
- Coseriu, Eugenio (1970): *Sprache – Strukturen und Funktionen*. Tübingen: Narr.
- Coseriu, Eugenio (1973): *Sincronía, diacronía e historia. El problema del cambio lingüístico*. Madrid: Gredos.
- Coseriu, Eugenio (1974): *Synchronie, Diachronie und Geschichte: Das Problem des Sprachwandels*. München: Fink.
- Cresti, Emanuela/ Moneglia, Massimo (Hrsg.) (2005): *C-ORAL-ROM. Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam [u.a.]: Benjamins.
- Cruz Piñol, Mar (2012): *Lingüística de corpus y enseñanza del español como 2/L*. Madrid: Arco Libros.
- Davies, Mark (2000): „Using multi-million word corpora of historical and dialectal Spanish texts to teach advanced courses in Spanish linguistics“. In: Burnard, L./ McEnery, T. (Hrsg.) (2000): *Rethinking language pedagogy from a corpus perspective: papers from the Third International Conference on Teaching and Language Corpora*. Frankfurt am Main: Peter Lang, S. 173-185.
- Davies, Mark (2004): „Student use of large, annotated corpora to analyze syntactic variation“. In: Aston, G./ Bernardini, S./ Stewart, D. (Hrsg.) (2004): *Corpora and Language Learners*. Amsterdam [u.a.]: Benjamins, S. 259-269.
- Davies, Mark (2005): „Vocabulary Range and Text Coverage: Insights from the Forthcoming *Routledge Frequency Dictionary of Spanish*“. In: Eddington, D. (Hrsg.) (2005): *Selected Proceedings of the 7th Hispanic Linguistics Symposium*. Somerville, MA: Cascadilla Proceedings Project, S. 106-116.
- Davies, Mark/ Face, Timothy L. (2006): „Vocabulary Coverage in Spanish Textbooks: How Representative is it?“. In: Sagarra, N./ Toribio, A. J. (Hrsg.): *Selected Proceedings of the 9th Hispanic Linguistics Symposium*. Somerville, MA: Cascadilla Proceedings Project, S. 132-143.
- De Alba Quiñones, Virginia (2012): *Lexías simples y colocaciones léxicas en la enseñanza de ELE. Análisis e implicaciones didácticas*. Frankfurt am Main: Peter Lang.
- De Florio-Hansen, Inez (2010): „Authentizität und Kompetenzorientierung beim Lernen und Lehren von Fremdsprachen“. In: Frings, M./ Leitzke-Ungerer, E. (Hrsg.)

- (2010): *Authentizität im Unterricht romanischer Sprachen*. Stuttgart: ibidem-Verlag, S. 263-282.
- De Jacobi, Claudia (2002): „Computadores, corpora y la enseñanza de español en cursos de letras“. In: *Anuario brasileño de estudios hispánicos*, 12, S. 29-44.
- Decke-Cornill, Helene (2004): „Die Kategorie der Authentizität im mediendidaktischen Diskurs der Fremdsprachendidaktik“. In: Bosenius, P./ Donnerstag, J. (Hrsg.) (2004): *Interaktive Medien und Fremdsprachenlernen*. Frankfurt am Main: Lang, S. 17-27.
- Díaz-Negrillo, Ana/ Ballier, Nicolas/ Thompson, Paul (Hrsg.) (2013): *Automatic Treatment and Analysis of Learner Corpus Data*. Amsterdam/ Philadelphia: John Benjamins.
- Doyé, Peter (1971): *Systematische Wortschatzvermittlung im Englischunterricht*. Hannover: Schroedel.
- Edelhoff, Christoph (1985): „Authentizität im Fremdsprachenunterricht“. In: Edelhoff, C. (Hrsg.) (1985): *Authentische Texte im Deutschunterricht*. München: Hueber, S. 7-30.
- Eisenmann, Maria (2016): „Binnendifferenzierung“. In: Burwitz-Melzer, E./ Mehlhorn, G./ Riemer, C./ Bausch, K.-R./ Krumm, H.-J. (Hrsg.) (2016): *Handbuch Fremdsprachenunterricht*. Tübingen: A. Francke Verlag, S. 358-361.
- Endruschat, Annette/ Ferreira, Vera (2011): „Dokumentationslinguistik – Ein Forschungsüberblick“. In: Endruschat, A./ Ferreira, V. (Hrsg.) (2011): *Sprachdokumentation und Korpuslinguistik – Forschungsstand und Anwendung: Akten des 8. Deutschen Lusitanistentags*. München: Meidenbauer, S. 1-30.
- Endruschat, Annette/ Schmidt-Radefeldt, Jürgen (2008): *Einführung in die portugiesische Sprachwissenschaft*. Tübingen: Narr Francke Attempto, S. 153.
- Enrique-Arias, Andrés (Hrsg.) (2009): *Diacronía de las lenguas iberorrománicas: nuevas aportaciones desde la lingüística de corpus*. Madrid: Iberorrománica.
- Eroms, Hans-Werner (2000): *Syntax der deutschen Sprache*. Berlin: De Gruyter.
- Escárate López, Luis M./ Steveker, Wolfgang/ Vences, Ursula/ Wlasak-Feik, Christine (2014): *Punto de vista. Nueva Edición. Texte und Aufgaben für die Oberstufe*. Berlin: Cornelsen.
- Fäcke, Christiane (2011): *Fachdidaktik Spanisch. Eine Einführung*. Tübingen: Narr Francke Attempto.

6. Bibliographie

- Felder, Ekkehard (2003): „Das Spannungsverhältnis zwischen Sprachnorm und Sprachvariation als Beitrag zu Sprach(differenz)bewusstheit“. In: *Wirkendes Wort*, 53. Jg., Heft 3/2003, S. 473–499.
- Fernández Pérez, Milagros (2011): *Lingüística de corpus y adquisición de la lengua*. Madrid: Arco Libros.
- Fillmore, Charles J. (1992): „'Corpus linguistics' or 'Computer aided armchair linguistics'“. In: Svartvik, J. (Hrsg.) (1992): *Directions in corpus linguistics. Proceedings of Nobel Symposium 82, Stockholm, 4 - 8 August 1991*. Berlin [u.a.]: Mouton de Gruyter, S. 35-60.
- Firth, John Rupert (1957): *Papers in Linguistics 1934-1951*. London [u.a.]: Oxford University Press.
- Fligelstone, Steven (1993): „Some reflections on the question of teaching, from a corpus linguistics perspective“. In: *ICAME Journal. Computers in English Linguistics*. No. 17. April 1993. Bergen: ICAME, S. 97-109.
- Flowerdew, Lynn (2000): „Investigating referential and pragmatic errors in a learner corpus“. In: Burnard, L./ McEnery, T. (Hrsg.) (2000): *Rethinking language pedagogy from a corpus perspective: papers from the Third International Conference on Teaching and Language Corpora*. Frankfurt am Main: Lang, S. 145-154.
- Gamallo Otero, Pablo/ Garcia, Marcos/ del Río, Iria/ González López, Isaac (2015): „Avalingua. Natural language processing for automatic error detection“. In: Callies, M./ Götz, S. (Hrsg.) (2015): *Learner Corpora in Language Testing and Assessment*. Amsterdam: John Benjamins, S. 35-58.
- Gaskell, Delian/ Cobb, Tom (2004): „Can learners use concordance feedback for writing errors?“. In: *System*, 32, S. 301–319.
- Gauger, Hans-Martin (1972): *Zum Problem der Synonyme*. Tübingen: Präzis.
- Gerstenberg, Annette (2013): *Arbeitstechniken für Romanisten. Eine Anleitung für den Bereich Linguistik*. 2. Auflage. Berlin/ Boston: De Gruyter.
- Gilmore, Alex (2007): „Authentic materials and authenticity in foreign language teaching“. In: *Language Teaching* 40/2, S. 97-118.
- Goethals, Michaël (2003): „E.E.T.: the European English Teaching vocabulary-list“. In: Lewandowska-Tomaszczyk, B. (Hrsg.) (2003): *Practical Applications in language and computers*. Frankfurt: Peter Lang, S. 417–427.
- Gómez Molina, José Ramón (2004): „Los contenidos léxico-semánticos“. In: Sánchez Lobato, J./ Santos Gargallo, I. (Hrsg.) (2004): *Vademécum para la formación de*

profesores. Enseñar español como segunda lengua (L2)/ lengua extranjera (LE). Alcobendas: Sociedad General Española de Librería, S. 789-810.

Granger, Sylviane (2002): „A bird’s-eye view of learner corpus research“. In: Granger, S./ Hung, J./ Petch-Tyson, S. (Hrsg.) (2002): *Computer learner corpora, second language acquisition and foreign language teaching*. Amsterdam: John Benjamins, S. 3-33.

Greenbaum, Sidney (1970): *Verb-Intensifier collocations in English. An experimental approach*. The Hague: Mouton.

Grünewald, Andreas/ Küster, Lutz (2009): *Fachdidaktik Spanisch. Tradition, Innovation, Praxis*. Stuttgart: Klett.

Haase, Ingrid (1995): *Konditionalsätze im authentischen englischen Sprachgebrauch und im Englischunterricht: eine empirische Untersuchung*. Freie Universität Berlin, Dissertation.

Hahn, Angela (2000): „Grammar at its best: The development of a rule- and corpus-based grammar of English tenses“. In: Burnard, L./ McEnery, T. (Hrsg.) (2000): *Rethinking language pedagogy from a corpus perspective: papers from the Third International Conference on Teaching and Language Corpora*. Frankfurt am Main: Peter Lang, S. 193-205.

Halliday, Michael Alexander Kirkwood (1993): „Quantitative studies and probabilities in grammar“. In: Hoey, M. (Hrsg.) (1993): *Data, Description, Discourse: Papers on the English Language in honour of John McH. Sinclair*. London: HarperCollins, S. 1-25.

Haß, Frank (2013): „Methoden im Fremdsprachenunterricht“. In: Hallet, W./ Königs, F. (Hrsg.) (2013): *Handbuch Fremdsprachendidaktik*. Seelze-Velber: Kallmeyer [u.a.], S. 151-156.

Hausmann, Franz Josef (1984): „Wortschatzlernen ist Kollokationslernen. Zum Lehren und Lernen französischer Wortverbindungen“. In: *Praxis des neusprachlichen Unterrichts* 1984, Bd. 31, S. 395-406.

Hausmann, Franz Josef (1985): „Kollokationen im deutschen Wörterbuch. Ein Beitrag zur Theorie des lexikographischen Beispiels“. In: Bergenholtz, H./ Mugdan, J. (Hrsg.) (1985): *Lexikographie und Grammatik. Akten des Essener Kolloquiums zur Grammatik im Wörterbuch 28.–30.06.1984*. Tübingen: Niemeyer, S. 118-129.

Hausmann, Franz Josef (2004): „Was sind eigentlich Kollokationen“. In: Steyer, K. (2004) (Hrsg.): *Wortverbindungen – mehr oder weniger fest*. Berlin: De Gruyter, S. 309-334.

Heger, Klaus (1970): „Belegbarkeit, Akzeptabilität und Häufigkeit. Zur Aufgabenstellung der Sprachwissenschaft“. In: Pilch, H./Richter, H. (Hrsg.)

6. Bibliographie

- (1970): *Theorie und Empirie in der Sprachforschung*. Basel/ München/ New York: Karger, S. 23-33.
- Hennemann, Anja (2013): „*Ayer he hecho* vs. *ayer yo hice*. On the increasing use of the (Peninsular) Spanish Present Perfect to describe completed actions“. In: Hennemann, A./ Schlaak, C. (Hrsg.) (2013): *Korpuslinguistische Untersuchungen: Analysen einzelsprachlicher Phänomene*. Berlin: Frank & Timme, S. 101-118.
- Heringer, Hans Jürgen (2009): „Ist das Deutsche grammatisch zu erfassen?“. In: Konopka, M./ Strecker, B. (Hrsg.) (2009): *Deutsche Grammatik – Regeln, Normen, Sprachgebrauch*. Berlin: De Gruyter, S. 23-32.
- Higgins, Derrick/ Ramineni, Chaitanya/ Zechner, Klaus (2015): „Learner Corpora and Automated Scoring“. In: Granger, S./ Gilquin, G./ Meunier, F. (Hrsg.) (2015): *The Cambridge Handbook of Learner Corpus Research*. Cambridge: Cambridge University Press, S. 587-604.
- Hinkel, Eli (2004): „Tense, aspect and the passive voice in L1 and L2 academic texts“. In: *Language Teaching Research* 8(1), S. 5–29.
- Hoey, Michael (2000): „The hidden lexical clues of textual organisation: a preliminary investigation into an unusual text from corpus perspective“. In: Burnard, L./ McEnery, T. (Hrsg.) (2000): *Rethinking language pedagogy from a corpus perspective: papers from the Third International Conference on Teaching and Language Corpora*. Frankfurt am Main: Peter Lang, S. 31-41.
- Holec, Henry (1981): *Autonomy and Foreign Language Learning*. Oxford [u.a.]: Pergamon Press.
- Hümmer, Christiane (2009): *Synonymie bei phraseologischen Einheiten: eine korpusbasierte Untersuchung*. Frankfurt am Main [u.a.]: Lang.
- Hundt, Marianne/ Nesselhauf, Nadja/ Biewer, Carolin (2007): *Corpus Linguistics and the Web*. Amsterdam and New York: Rodopi.
- Hundt, Markus (2008): „Grammatikalität – Akzeptabilität – Sprachnorm. Zum Verhältnis von Korpuslinguistik und Grammatikalitätsurteilen“. In: Lenz, F./ Schierholz, S. (Hrsg.) (2008): *Korpuslinguistik in Lexik und Grammatik*. Tübingen: Stauffenburg, S. 15-40.
- Hundt, Markus (2010): „New norms – How new grammatical constructions emerge“. In: Lenz, A. N./ Plewnia, A. (Hrsg.) (2010): *Grammar between Norm and Variation*. Frankfurt: Lang, S. 27-57.
- Hunston, Susan (2002): *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

- Iliescu, Maria (2006): „Kollokationen in den romanischen Sprachen“. In: Dietrich, W./ Hoinkes, U./ Roviró, B./ Warnecke, M. (Hrsg.) (2006): *Lexikalische Semantik und Korpuslinguistik*. Tübingen: Narr Francke Attempto Verlag, S. 191-208.
- Irsula Peña, Jesús (1994): *Substantiv-Verb-Kollokationen: Kontrastive Untersuchungen Deutsch-Spanisch*. Frankfurt am Main: Peter Lang.
- James, Carl (1996): „A cross-linguistic approach to language awareness“. In: *Language Awareness* 5, S. 138-148.
- James, Carl/ Garrett, Peter (1992): *Language awareness in the classroom. Applied linguistics and language study*. London: Longman.
- Johns, Tim (1991): „Should you be persuaded’: Two samples of data-driven learning materials“. In: Johns, T./ King, P. (Hrsg.) (1991): *Classroom Concordancing ELR Journal*, 4. University of Birmingham, S. 1–16.
- Johns, Tim (1997): „Contexts: The background, development and trialling of a concordance-based CALL program“. In: Wichmann, A./ Fligelstone, S./ McEnery, T./ Knowles, G. (Hrsg.) (1997): *Teaching and language corpora*. Harlow: Longman, S. 100–115.
- Jones, Roger D./ Stuhlmann, Sebastian/ Zeyer, Tamara (2016): „Interaktives Fremdsprachenlernen: Potenziale und Herausforderungen“. In: Zeyer, T./ Stuhlmann, S./ Jones R. (Hrsg.) (2016): *Interaktivität beim Fremdsprachenlehren und –lernen mit digitalen Medien. Hit oder Hype?* Tübingen: Narr Francke Attempto, S. 11-42.
- Kabatek, Johannes (Hrsg.) (2016): *Lingüística de corpus y lingüística histórica iberorrománica*. Berlin/ Boston: Walter De Gruyter.
- Kabatek, Johannes/ Pusch, Claus D. (2011): *Spanische Sprachwissenschaft*. Tübingen: Narr Francke Attempto.
- Kaltenböck, Gunther/ Mehlmauer-Larcher, Barbara (2005): „Computer corpora and the language classroom: On the potential and limitations of computer corpora in language teaching“. In: *ReCALL*, 17, S. 65-84.
- Kaszubski, Przemyslaw/ Wojnowska, Aleksandra (2003): „Corpus-informed exercises for learners of English: The TestBuilder program“. In: Oleksy, E./ Lewandowska-Tomaszczyk, B. (Hrsg.) (2003): *Research and scholarship in integration processes: Poland–USA–EU*. Łódź: Łódź University Press, S. 337–354.
- Keck, Casey Mari (2004): „Corpus linguistics and language teaching research: Bridging the gap“. In: *Language Teaching Research*, 8 (1), S. 83-109.
- Kennedy, Graeme (1998): *An introduction to corpus linguistics*. Harlow: Longman.

6. Bibliographie

- Kilgarriff, Adam (2006): „Googleology is Bad Science“. *Computational Linguistics* 1/33, S. 147-151.
- Kilgarriff, Adam/ Grefenstette, Gregory (2003): „Web as Corpus“. In: *Computational Linguistics* 3/29. Online Publikation: <https://www.kilgarriff.co.uk/Publications/2003-KilgGrefenstette-WACIntro.pdf> (11.11.2016).
- Kispál, Tamás (2007): „Sprichwörtersammlungen“. In: Burger, H./ Dobrovol'skij, D./ Kühn, P./ Norrick, N. R. (2007): *Phraseologie. Ein internationales Handbuch der zeitgenössischen Forschung*. HSK Band 28.1. Berlin [u.a.]: De Gruyter, S. 414-423.
- Klöden, Hildegard (2002): „Romanistische Korpuslinguistik: Leistungen, Grenzen, Perspektiven“. In: Pusch, C./ Raible, W. (Hrsg.) (2002): *Romanistische Korpuslinguistik. Romance Corpus Linguistics*. Korpora und gesprochene Sprache. Corpora and Spoken Language (= ScriptOralia 126). Tübingen: Narr, S. 7-18.
- Kloudová, Věra (2015): *Synonymie und Antonymie*. Heidelberg: Universitätsverlag Winter.
- Koch, Peter/ Oesterreicher, Wulf (2011): *Gesprochene Sprache in der Romania. Französisch, Italienisch, Spanisch*. Berlin/ New York: De Gruyter.
- Koike, Kazumi (2001): *Colocaciones léxicas en el español actual: Estudio formal y léxico-semántico*. Alcalá de Henares: Universidad de Alcalá de Henares [u.a.].
- Kramsch, Claire J. (1993): *Context and culture in language teaching*. Oxford: Oxford Univ. Press.
- Krashen, Stephen D. (1996): „Not pedagogical or authentic, but interesting and comprehensible...“. In: *Zielsprache Englisch*, 1996, 26 (2), S. 22-25.
- Kupietz, Marc/ Keibel, Holger (2009): „Gebrauchsbasierte Grammatik: Statistische Regelmäßigkeit“. In: Konopka, M./ Strecker, B. (Hrsg.) (2009): *Deutsche Grammatik – Regeln, Normen, Sprachgebrauch*. Berlin: De Gruyter, S. 33-50.
- Lademann, Norbert (1997): „Authentizität, Text und Sprechintention“. In: Siebold, J. (Hrsg.) (1997): *Sprache und Medien im Fremdsprachenunterricht. Beiträge des 3. Mediendidaktischen Kolloquiums Oktober 1996 in Rostock*. Rostock: Universitätsdruckerei Rostock, S. 39-48.
- Lahuerta, Javier/ Pujol, Mercè (1996): „El lexicón mental y la enseñanza del vocabulario“. In: Segoviano, C. (Hrsg.) (1996): *La enseñanza del léxico español como lengua extranjera*. Madrid: Iberoamericana, S. 117-129.
- Leech, Geoffrey (1992): „Corpora and theories of linguistic performance“. In: Svartvic, J. (Hrsg.) (1992): *Directions in corpus linguistics. Proceedings of Nobel*

- Symposium 82, Stockholm, 4-8 August, 1991*. Berlin [u.a.]: Mouton de Gruyter, S. 105-122.
- Leech, Geoffrey (1997): „Teaching and Language Corpora: a Convergence.“ In: Wichmann, A./ Fligelstone, S./ McEnery, T./ Knowles, G. (Hrsg.) (1997): *Teaching and Language Corpora*. London: Longman, S. 1-23.
- Leech, Geoffrey (2007): „New resources, or just better old ones? The Holy Grail of representativeness“. In: Hundt, M./ Nesselhauf, N./ Biewer, C. (Hrsg.) (2007): *Corpus Linguistics and the Web*. Amsterdam/ New York: Rodopi, S. 133-149.
- Leech, Geoffrey/ Fligelstone, Steven (1992): „Computers and Corpus Analysis“. In: Butler, C. (Hrsg.) (1992): *Computers and Written Texts*. Oxford: Blackwell, S. 115-140.
- Leitzke-Ungerer, Eva (2010): „Zielkulturelle und lernkontextbezogene Authentizität im Fremdsprachenunterricht“. In: Frings, M./ Leitzke-Ungerer, E. (Hrsg.) (2010): *Authentizität im Unterricht romanischer Sprachen*. Stuttgart: ibidem-Verlag, S. 11-24.
- Lemnitzer, Lothar/ Zinsmeister, Heike (2010): *Korpuslinguistik. Eine Einführung*. Tübingen: Narr Francke Attempto Verlag.
- Leńko-Szymańska, Agnieszka (2003): „Lexical problem areas in the advanced learner corpus of written data“. In: Lewandowska-Tomaszczyk, B./ Melia, P.J. (Hrsg.) (2003): *PALC 2001: Practical Applications in Language Corpora*. Frankfurt am Main: Peter Lang, S. 505-519.
- Leńko-Szymańska, Agnieszka (2004): „Demonstratives as anaphora markers in advanced learners' English“. In: Aston, G./ Bernardini, S./ Stewart, D. (Hrsg.) (2004): *Corpora and Language Learners*. Amsterdam [u.a.]: Benjamins, S. 89-108.
- Little, David/ Devitt, Seán/ Singleton, David M. (1994): *Learning foreign languages from authentic texts. Theory and practice*. Dublin: Authentik Language Learning Resources.
- Lloyd, Paul M. (1987): *From Latin to Spanish. Vol. I: Historical Phonology and Morphology of the Spanish Language*. Philadelphia: American Philosophical Society.
- López-Ferrero, Carmen (2007): „Academic writing: Exploring Corpus 92“. In: Parodi, G. (Hrsg.) (2007): *Working With Spanish Corpora*. London: Continuum, S. 173-194.
- Luque Toro, Luis (2006): „El concepto de colocación y su presencia en los diccionarios de uso del español actual“. In: Alonso Ramos, M. (Hrsg.) (2006): *Diccionarios y Fraseología*. A Coruña: Servizo de Publ., Univ. da Coruña, S. 123-131.

6. Bibliographie

- Marková, Věra (2012): *Synonyme unter dem Mikroskop: eine korpuslinguistische Studie*. Tübingen: Narr.
- McAlpine, Janice/ Myles, Johanne (2003): „Capturing phraseology in an online dictionary for advanced users of English as a second language: A response to user needs“. In: *System*, 31, S. 71-84.
- McCullough, James L. (2001): „Los usos de los corpórea de textos en la enseñanza de lenguas“. In: Trenchs Parera, M. (Hrsg.) (2001): *Nuevas Tecnologías para el Autoaprendizaje y la Didáctica de Lenguas*. Lleida: Milenio, S. 125-140.
- McEnery, Tony/ Hardie, Andrew (2012): *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.
- McEnery, Tony/ Wilson, Andrew (1996): *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- McEnery, Tony / Xiao, Richard (2010): „What Corpora Can Offer in Language Teaching and Learning“. In: Hinkel, E. (Hrsg.) (2010): *Handbook of Research in Second Language Teaching and Learning: Volume 2*. London/ New York: Routledge, S. 364-380.
- Meißner, Franz-Joseph (2006): „Linguistische und didaktische Überlegungen zur Entwicklung von Kompetenzaufgaben im Lernbereich Mündlichkeit“. In: *Französisch heute* 3, S. 240-283.
- Mellado Blanco, Carmen (2008): „Introducción: colocaciones y algunas cuestiones teórico-prácticas de fraseografía“. In: Mellado Blanco, C. (2008): *Colocaciones y fraseología en los diccionarios*. Frankfurt am Main: Peter Lang, S. 7-31.
- Meunier, Fanny (2002): „The pedagogical value of native and learner corpora in EFL grammar teaching“. In: Granger, S./ Hung, J./ Petch-Tyson, S. (Hrsg.) (2002): *Computer learner corpora, second language acquisition and foreign language teaching*. Philadelphia: John Benjamins, S. 119-142.
- Mindt, Dieter (1996): „English corpus linguistics and the foreign language teaching syllabus“. In: Thomas, J./ Short, M. (Hrsg.) (1996): *Using Corpora for Language Research*. London: Longman, S. 232-247.
- Montoro del Arco, Esteban Tomás (2010): „Norma y uso de las colocaciones léxicas del español“. In: Aleza Izquierdo, M. (2010): *Normas y usos correctos en el español actual*. Valencia: Tirant Humanidades, S. 253-269.
- Moreno Fernández, Francisco (2002): *La lengua hablada en Alcalá de Henares. Corpus PRESEEA-Alcalá*. Alcalá: Univ. de Alcalá. CD-ROM.
- Müller, Simone (2005): *Discourse markers in native and non-native English discourse*. Amsterdam: John Benjamins.

- Mukherjee, Joybrato/ Rohrbach, Jan-Marc (2006): „Rethinking applied corpus linguistics from a language-pedagogical perspective: New departures in learner corpus research“. In: Kettemann, B./ Marko, G. (Hrsg.) (2006): *Planning, gluing and painting corpora: Inside the applied corpus linguist's workshop*. Frankfurt: Peter Lang, S. 205–232.
- Nelson, Michael (2000): *A corpus-based study of business English and business English teaching materials*. PhD thesis. University of Manchester. Online Publikation: <http://users.utu.fi/micnel/thesis.html> (19.07.2017).
- Nesselhauf, Nadja (2004): „How learner corpus analysis can contribute to language teaching: A study of support verb constructions“. In: Aston, G./ Bernardini, S./ Stewart, D. (Hrsg.) (2004): *Corpora and Language Learners*. Amsterdam [u.a.]: Benjamins, S. 109-124.
- Nicolás Martínez, Carlota (2012): *Corpus C-Or-DiAL (Corpus Oral Didáctico Anotado Lingüísticamente)*. Madrid: Liceus.
- O’Keeffe, Anne/ Farr, Fiona (2003): „Using language corpora in initial teacher education: pedagogic issues and practical applications“. In: *TESOL Quarterly*, 37 (3), S. 389–418.
- Osborne, John (2000): „What can students learn from a corpus?: building bridges between data and explanation“. In: Burnard, L./ McEnery, T. (Hrsg.) (2000): *Rethinking language pedagogy from a corpus perspective: papers from the Third International Conference on Teaching and Language Corpora*. Frankfurt am Main: Peter Lang, S. 165-172.
- Osborne, John (2003): „Integrating corpora into a language-learning syllabus“. In: Lewandowska-Tomaszczyk, B. (Hrsg.) (2003): *PALC 2001: Practical Applications in Language Corpora*. Frankfurt am Main: Lang, S. 478-492.
- Osborne, John (2004): „Top-down and Bottom-up Approaches to Corpora in Language Teaching“. In: Connor, U./ Upton, T. A. (Hrsg.) (2004): *Applied Corpus Linguistics. A Multidimensional Perspective*. Amsterdam [u.a.]: Rodopi, S. 251-266.
- O’Sullivan, Íde/ Chambers, Angela (2006): „Learners’ writing skills in French: Corpus consultation and learner evaluation“. In: *Journal of Second Language Writing*, 15, S. 49-68.
- Padrós, Alicia/ Biechele, Markus (2003): *Didaktik der Landeskunde*. Berlin/ München: Langenscheidt.
- Parodi, Giovanni (Hrsg.) (2007): *Working With Spanish Corpora*. London: Continuum.
- Parodi, Giovanni (2007a): „Introduction. Catching up with corpus linguistics: Register-

6. Bibliographie

- diversified studies from different corpora in different Spanish-speaking countries". In: Parodi, G. (Hrsg.) (2007): *Working With Spanish Corpora*. London: Continuum, S. 1-10.
- Parodi, Giovanni (2007b): „Variation across registers in Spanish: Exploring the El Grial PUCV Corpus". In: Parodi, G. (Hrsg.) (2007): *Working With Spanish Corpora*. London: Continuum, S. 11-53.
- Parodi, Giovanni (2010): *Lingüística de corpus: de la teoría a la empiria*. Madrid: Iberoamericana.
- Parodi, Giovanni/ Gramajo, Aída (2007): „Technical-professional discourses: Specialized and dissemination text types". In: Parodi, G. (Hrsg.) (2007): *Working With Spanish Corpora*. London: Continuum, S. 145-172.
- Pearse, Roy (1983): „Realistic and Real English in the Classroom". In: *English Teaching Forum*, 21/3, S. 19-22.
- Petry, Kristina (2015): *RAAbits Spanisch. Grundwerk*. Stuttgart: Dr. Josef Raabe Verlag, M6/M12.
- Piepho, Hans-Eberhard (1974): *Kommunikative Kompetenz, Pragmalinguistik und Ansätze zur Neubestimmung in der Lernzielbestimmung im Fremdsprachenunterricht*. Düsseldorf: Winterscheidt.
- Pöll, Bernhard (2002): *Spanische Lexikologie. Eine Einführung*. Tübingen: Narr.
- Quinteiro, Natalia/ Quiroz, Gabriel (2012): „Analysis of the Translation of the Phraseological Units of Cien Años de Soledad into English: A Corpus-based Study". In: Pezik, P. (Hrsg.) (2012): *Corpus Data across Languages and Disciplines*. Frankfurt am Main [u.a.]: Lang, S. 245-254.
- Rautenhaus, Heike (1997): „Authentische Texte und Konkordanzprogramme im Englischunterricht". In: Siebold, J. (Hrsg.) (1997): *Sprache und Medien im Fremdsprachenunterricht. Beiträge des 3. Mediendidaktischen Kolloquiums Oktober 1996 in Rostock*. Rostock: Universitätsdruckerei Rostock, S. 141-166.
- Rivas, Javier/ Brown, Esther (2010): „Variable development of intersubjectivity in Spanish". In: Sánchez, A./ Almela, M. (Hrsg.) (2010): *A Mosaic of Corpus Linguistics: Selected Approaches*. Frankfurt am Main: Lang, S. 61-78.
- Römer, Ute (2004): „Comparing real and ideal language learner input: The use of an EFL textbook corpus in corpus linguistics and language teaching". In: Aston, G./ Bernardini, S./ Stewart, D. (Hrsg.) (2004): *Corpora and Language Learners*. Amsterdam [u.a.]: John Benjamins, S. 151-168.

- Römer, Ute (2005): *Progressives, patterns, pedagogy. A corpus-driven approach to English progressive forms, functions, contexts and didactics*. Amsterdam: John Benjamins.
- Rössler, Andrea (2010): „Zur Frage der Authentizität von Lehrwerkdialogen für den Französischunterricht“. In: Frings, M./ Leitzke-Ungerer, E. (Hrsg.) (2010): *Authentizität im Unterricht romanischer Sprachen*. Stuttgart: ibidem-Verlag, S. 27-44.
- Rüschhoff, Bernd (1997): „Multimedia Autorentools und Authentizität im Fremdsprachenlernen“. In: Siebold, J. (Hrsg.) (1997): *Sprache und Medien im Fremdsprachenunterricht. Beiträge des 3. Mediendidaktischen Kolloquiums Oktober 1996 in Rostock*. Rostock: Universitätsdruckerei Rostock, S. 63-82.
- Rüschhoff, Bernd/ Wolff, Dieter (1999): *Fremdsprachenlernen in der Wissensgesellschaft. Zum Einsatz der neuen Technologien in Schule und Unterricht*. Ismaning: Hueber.
- Samper Padilla, Antonio José (1998): *Macrocorpus de la norma lingüística culta de las principales ciudades del mundo hispánico. Intro*. Las Palmas de Gran Canaria: Univ. de las Palmas. CD-ROM.
- Sampson, Geoffrey (1996): „From central embedding to corpus linguistics“. In: Thomas, J./ Short, M. (Hrsg.) (1996): *Using Corpora for Language Research*. London: Longman, S. 14-26.
- Sánchez, Aquilino/ Cantos, Pascual/ Simón Granda, José (2001): *Corpus Cumbre lingüístico: selección de 2.000.000 de palabras del lenguaje oral y escrito*. Madrid: SGEL, CD-ROM.
- Schaeder, Burkhard (1976): „Maschinenlesbare Textkorpora des Deutschen und des Englischen“. In: *Deutsche Sprache*, 4, S. 356-370.
- Schank, Gerd/ Schoenthal, Gisela (1976): *Gesprochene Sprache. Eine Einführung in Forschungsansätze und Analysemethoden*. Tübingen: De Gruyter.
- Scherer, Carmen (2006): *Korpuslinguistik*. Heidelberg: Universitätsverlag Winter.
- Schiffler, Ludger (2012): *Effektiver Fremdsprachenunterricht*. Tübingen: Narr Francke Attempto.
- Schirn, Matthias (1975): *Identität und Synonymie. Logisch-semantische Untersuchungen unter Berücksichtigung der sprachlichen Verständigungspraxis*. Stuttgart/ Bad Cannstadt: Frommann/ Holzboog.
- Schlickau, Stephan (2009): *Neue Medien in der Sprach- und Kulturvermittlung*. Frankfurt am Main: Lang.

6. Bibliographie

- Schlüter, Norbert (2002): *Present Perfect. Eine korpuslinguistische Analyse des englischen Perfekts mit Vermittlungsvorschlägen für den Sprachunterricht*. Tübingen: Narr.
- Schmelter, Lars (2004): *Selbstgesteuertes und potenziell expansives Fremdsprachenlernen im Tandem*. Tübingen: Narr.
- Schmenk, Barbara (2016): „Lernerautonomie und selbst gesteuertes Sprachenlernen“. In: Burwitz-Melzer, E./ Mehlhorn, G./ Riemer, C./ Bausch, K.-R./ Krumm, H.-J. (Hrsg.) (2016): *Handbuch Fremdsprachenunterricht*. Tübingen: A. Francke Verlag, S. 367-372.
- Schmidt, Richard (1994): „Deconstructing consciousness in search of useful definitions for applied linguistics“. In: *AILA Review*, 11, S. 11-26.
- Schütz, Susanne (2013): „Un paquete de Buenos Aires. Mehrkanaliges Lernen an landeskundlichen Stationen“. In: *Der Fremdsprachliche Unterricht Spanisch – Metrópolis*, 2013, Band 40, S. 30-35, inklusive CD-ROM.
- Sedano, Mercedes (2007): „Future tense expressions in several Spanish corpora“. In: Parodi, G. (Hrsg.) (2007): *Working With Spanish Corpora*. London: Continuum, S. 132-144.
- Seidlhofer, Barbara (2000): „Operationalizing intertextuality: using learner corpora for learning“. In: Burnard, L./ McEnery, T. (Hrsg.) (2000): *Rethinking language pedagogy from a corpus perspective: papers from the Third International Conference on Teaching and Language Corpora*. Frankfurt am Main: Lang, S. 207-223.
- Selinker, Larry (1972): „Interlanguage“. In: *International Review of Applied Linguistics in Language Teaching*, 10 (3), S. 209-231.
- Siepmann, Dirk (2009): „Was der Fremdsprachenunterricht von der Korpuslinguistik erwarten darf“. In: Jung, U. O. H. (Hrsg.) (2009): *Praktische Handreichung für Fremdsprachenlehrer*. Lang: Frankfurt am Main [u.a.], S. 321-330.
- Siepmann, Dirk (2015): „Dictionaries and spoken language: a corpus-based review of French dictionaries“. In: *International Journal of Lexicography*. Band 28, Ausgabe 2, 1. Juni 2015, S. 139-168.
- Siepmann, Dirk/ Bürgel, Christoph/ Diwersy, Sascha (2016): „Le Corpus de référence du français contemporain (CRFC), un corpus massif du français largement diversifié par genres“. In: *Congrès Mondial de Linguistique Française – CMLF 2016*.
- Siepmann, Dirk/ Bürgel, Christoph/ Diwersy, Sascha (2017): „The Corpus de référence du français contemporain (CRFC) as the First Genre-Diverse Mega-Corpus of French“. In: *International Journal of Lexicography*, Volume 30, Issue 1, S. 63-84.

- Simpson, Rita/ Lucka, Bret/ Ovens, Janine (2000): „Methodological challenges of planning a spoken corpus with pedagogical outcomes“. In: Burnard, L./ McEnery, T. (Hrsg.) (2000): *Rethinking language pedagogy from a corpus perspective: papers from the Third International Conference on Teaching and Language Corpora*. Frankfurt am Main: Peter Lang, S. 43-49.
- Sinclair, John McHardy (1987): „Collocation: a progress report“. In Steele, R./ Threadgold, T. (Hrsg.) (1987): *Language topics: essays in honour of Michael Halliday*. Amsterdam: John Benjamins, S. 319-331.
- Sinclair, John McHardy (1991): *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, John McHardy (Hrsg.) (2004): *How to Use Corpora in Language Teaching*. Amsterdam [u.a.]: Benjamins.
- Sinclair, John McHardy/ Renouf, Antoinette (1988): „A lexical syllabus for language learning“. In: Carter, R. / McCarthy, M. (Hrsg.) (1988): *Vocabulary and language teaching*. Harlow: Longman, S. 140–158.
- Sonntag, Eric (2007): „Diminutivische Sachbezeichnungen. Lexikalisierte Formen auf –illa zwischen Lexikographie und Korpuslinguistik“. In: Bierbach, M./ von Gemmingen, B./ Stork, Y. (Hrsg.) (2007): *Das gefesselte Wort. Beiträge zur Entwicklung von Wörterbüchern und Grammatiken des Spanischen*. Bonn: Romanistischer Verlag, S. 183-194.
- Sripicharn, Passapong (2004): „Examining native speakers’ and learners’ investigation of the same concordance data and its implications for classroom concordancing with ELF learners“. In: Aston, G./ Bernardini, S./ Stewart, D. (Hrsg.) (2004): *Corpora and Language Learners*. Amsterdam [u.a.]: Benjamins, S. 233-244.
- Stefanowitsch, Anatol (2005): „Quantitative Korpuslinguistik und sprachliche Wirklichkeit“. In Solte-Gresser, C./ Struve, K./ Ueckmann, N. (Hrsg.) (2005): *Von der Wirklichkeit zur Wissenschaft. Aktuelle Forschungsmethoden in den Sprach-, Literatur- und Kulturwissenschaften*. Münster: LIT Verlag, S. 147-161.
- Stubbs, Michael (1995): „Collocations and semantic profiles. On the cause of the trouble with quantitative studies“. In: *Functions of Language* 2, 1 (1995). Online Publikation: <https://www.uni-trier.de/fileadmin/fb2/ANG/Linguistik/Stubbs/stubbs-1995-cause-trouble.pdf> (30.09.2017).
- Stubbs, Michael (2000): 'Society, education and language: the last 2,000 (and the next 20?) years of language teaching'. In: Trappes-Lomax, H. (Hrsg.) (2000): *Change and Continuity in Applied Linguistics*. Clevedon: BAAL and Multilingual Matters, S. 15-34. Online Publikation: <https://www.uni-trier.de/fileadmin/fb2/ANG/Linguistik/Stubbs/stubbs-2000-soc-ed-lg.pdf> (19.07.2017).

6. Bibliographie

- Summers, Della (1996): „Computer lexicography: the importance of representativeness in relation to frequency “. In Thomas, J./ Short, M. (Hrsg.) (1996): *Using Corpora for Language Research*. London: Longman, S. 260-266.
- Tagnin, Stella E.O. (2003): „Comet – A multilingual corpus for teaching and translation“. In: Lewandowska-Tomaszczyk, B. (Hrsg.) (2003): *PALC 2001: Practical Applications in Language Corpora*. Frankfurt am Main: Lang, S. 536-540.
- Tankó, Guyla (2004): „The use of adverbial connectors in Hungarian university students' argumentative essays“. In: Sinclair, J. McH. (Hrsg.) (2004): *How to Use Corpora in Language Teaching*. Amsterdam [u.a.]: Benjamins, S. 157-184.
- Taylor, Lynda/ Barker, Fiona (2008): „Using Corpora for Language Assessment“. In: Shohamy, E./ Hornberger N.H. (Hrsg.) (2008): *Encyclopedia of Language and Education. Language Testing and Assessment*. Second Edition, Vol. 7, New York: Springer, S. 241-254.
- Tesch, Felicitas (1990): *Die Indefinitpronomina „some“ und „any“ im authentischen englischen Sprachgebrauch und in Lehrwerken*. Tübingen: Narr.
- Teubert, Wolfgang (2007): „Language and corpus linguistics“. In: Teubert, W./ Čermáková, A. (Hrsg.) (2007): *Corpus Linguistics. A Short Introduction*. London: Continuum, S. 1-58.
- Teubert, Wolfgang/ Čermáková, Anna (2007): „Directions in corpus linguistics“. In: Teubert, W./ Čermáková, A. (Hrsg.) (2007): *Corpus Linguistics. A Short Introduction*. London: Continuum, S. 59-134.
- Thurstun, Jennifer/ Candlin, Christopher N. (1997): *Exploring academic English: A workbook for student essay writing*. Sydney: NCELTR.
- Thurstun, Jennifer/ Candlin, Christopher N. (1998): „Concordancing and the teaching of the vocabulary of academic English“. In: *English for Specific Purposes* 17, S. 267–280.
- Tognini-Bonelli, Elena (2001): *Corpus Linguistics at Work*. Amsterdam/ Philadelphia: John Benjamins.
- Tolchinsky, Liliana (2014): „El uso de corpus lingüísticos como herramienta pedagógica“. In: *Textos de Didáctica de la Lengua y de la Literatura*, enero 2014, núm. 65, S. 9-17.
- Tono, Yukio (2000): „A computer learner corpus based analysis of the acquisition order of English grammatical morphemes“. In: Burnard, L./ McEnery, T. (Hrsg.) (2000): *Rethinking language pedagogy from a corpus perspective: papers from the Third International Conference on Teaching and Language Corpora*. Frankfurt am Main: Lang, S. 123-132.

- Tracy-Ventura, Nicole/ Cortes, Viviana/ Biber, Douglas (2007): „Lexical bundles in speech and writing”. In: Parodi, G. (Hrsg.) (2007): *Working With Spanish Corpora*. London: Continuum, S. 217-231.
- Trenchs Parera, Mireia (2001): „Tecnologías digitales para un cambio educativo en el aprendizaje de lenguas“. In: Trenchs Parera, M. (Hrsg.) (2001): *Nuevas Tecnologías para el Autoaprendizaje y la Didáctica de Lenguas*. Lleida: Milenio, S. 21-34.
- Valverde-Ibáñez, Pilar M./ Bick, Eckhard (2010): „A web corpus of Spanish automatically annotated with semantic roles“. In: Sánchez, A./ Almela, M. (Hrsg.) (2010): *A Mosaic of Corpus Linguistics: Selected Approaches*. Frankfurt am Main: Lang, S. 249-268.
- Van Lier, Leo (1996): *Interaction in the Language Curriculum*. London: Longman.
- Vande Castele, An (2010): *Las construcciones apositivas de referente humano en el discurso periodístico informativo*. Tübingen: Narr Francke Attempto.
- Vázquez Veiga, Nancy (2016): „Discourse markers in CEDEL2 and SPLLOC corpora of learner Spanish. Analysis of some lexical-pragmatic failures“. In: Alonso-Ramos, M. (Hrsg.) (2016): *Spanish Learner Corpus Research: Current trends and future perspectives*. Amsterdam [u.a.]: John Benjamins, S. 267-297.
- Vega, Jorge (2013): „Towards a subject cartography in Catalan: the example of oblique subjects“. In: Hennemann, A./ Schlaak, C. (Hrsg.) (2013): *Korpuslinguistische Untersuchungen: Analysen einzelsprachlicher Phänomene*. Berlin: Frank & Timme, S. 89-100.
- Widdowson, Henry G. (1979): *Teaching Language as Communication*. Oxford: University Press.
- Willis, Dave (1990): *The lexical syllabus: A new approach to language teaching*. London: HarperCollins.
- Willis, Jane/ Willis, Dave (1989): *Collins COBUILD English Course*. London: HarperCollins.
- Yoon, Hyunsook/ Hirvela, Alan (2004): „ESL student attitudes toward corpus use in L2 writing“. In: *Journal of Second Language Writing*, 13, S. 257–283.
- Ziff, Paul (1960): *Semantic Analysis*. Ithaka, NY: Cornell University Press.
- Zulaica Hernández, Iker (2010): „Testing accessibility and givenness in a corpus of Spanish“. In: Sánchez, A./ Almela, M. (Hrsg.) (2010): *A Mosaic of Corpus Linguistics: Selected Approaches*. Frankfurt am Main: Lang, S. 99-118.

6.3. Internetquellen

Alonso Pérez-Ávila, Elena (2006): *El corpus lingüístico en la didáctica del léxico en el aula de E/LE*. Online Publikation: <https://www.mecd.gob.es/dam/jcr:c9948929-a65a-4afc-8df2-9c104e1e6175/2010-bv-11-01alonso-pdf.pdf> (06.06.2017).

Alonso Pérez-Ávila, Elena (2007): *El corpus lingüístico en la didáctica del léxico del español como LE*. Online Publikation: <https://aperto.unito.it/retrieve/handle/2318/57837/6964/EI%20corpus%20lingü%C3%A4stico%20en%20la%20didáctica%20del%20léxico%20del%20español%20como%20LE.pdf> (06.06.2017).

Alonso Ramos, Margarita: *Diccionario de colocaciones del Español*. Online Publikation: <http://www.dicesp.com/paginas> (20.10.2017).

Alonso Ramos, Margarita: *Diccionario de colocaciones del Español. ¿Cómo es el DiCE?*. Online Publikation: <http://www.dicesp.com/paginas/index/1> (20.10.2017).

Alonso Ramos, Margarita: *Diccionario de colocaciones del Español. Consulta avanzada*. Online Publikation: <http://www.dicesp.com/paginas/index/6> (20.10.2017).

Alonso Ramos, Margarita: *Diccionario de colocaciones del Español. Consulta general*. Online Publikation: <http://www.dicesp.com/consultageneral/lemas> (20.10.2017).

Alvarado Ortega, María Belén (2008): *Las fórmulas rutinarias en el español actual*. Tesis Doctorales, Universidad de Alicante. Online Publikation: <https://www.google.de/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&cad=rja&uact=8&ved=2ahUKEwjL97GXhoLjAhXNJVAKHZbLAI0QFjABegQIABAC&url=http%3A%2F%2Fwww.cervantesvirtual.com%2FdescargaPdf%2Fflas-formulas-rutinarias-en-el-espaol-actual%2F&usg=AOvVaw1HsmoPE083-C6tN5A71161> (24.06.2019).

Anthony, Laurence (2010): *AntConc 3.2.4 – Tutorial 0: Getting Started*. Online Publikation: <https://www.youtube.com/watch?v=z9wwX7eR-Y> (15.02.2015).

Anthony, Laurence (2011): *AntConc 3.2.4 – Tutorial 1: Concordance Tool – Basic Features*. Online Publikation: <https://www.youtube.com/watch?v=9TsqFVrUYO0> (16.02.2015).

Anthony, Laurence (2011): *AntConc 3.2.4 – Tutorial 2: Concordance Tool – Advanced Features*. Online Publikation: <https://www.youtube.com/watch?v=pUN-flv-Cmw> (17.02.2015).

Anthony, Laurence (2011): *AntConc 3.2.4 – Tutorial 3: Concordance Plot Tool – Basic Features*. Online Publikation: <https://www.youtube.com/watch?v=YNDvcxonmP4> (19.02.2015).

- Anthony, Laurence (2011): *AntConc 3.2.4 – Tutorial 4: File View Tool – Basic Features*. Online Publikation: https://www.youtube.com/watch?v=Yalioyw_74M (30.03.2015).
- Anthony, Laurence (2011): *AntConc 3.2.4 – Tutorial 5: Clusters Tool – Basic Features*. Online Publikation: <https://www.youtube.com/watch?v=5kosY5xlsLE> (01.04.2015).
- Anthony, Laurence (2011): *AntConc 3.2.4 – Tutorial 6: N-Grams Tool – Basic Features*. Online Publikation: <https://www.youtube.com/watch?v=-vv663HjTso> (02.04.2015).
- Anthony, Laurence (2011): *AntConc 3.2.4 – Tutorial 7: File Collocates Tool – Basic Features*. Online Publikation: <https://www.youtube.com/watch?v=An6gb1W3pFM> (02.04.2015).
- Anthony, Laurence (2011): *AntConc 3.2.4 – Tutorial 8: Word List Tool – Basic Features*. Online Publikation: https://www.youtube.com/watch?v=Zb71yaBP_II (07.04.2015).
- Anthony, Laurence (2011): *AntConc 3.2.4 Tutorial 9: Keyword List Tool - Basic Features*. Online Publikation: <https://www.youtube.com/watch?v=JvasTvQY7kU> (22.07.2015).
- Anthony, Laurence (2014): *AntConc 3.4.0 – Tutorial 6: Clusters Tool*. Online Publikation: <https://www.youtube.com/watch?v=DV8YC0zKBOg> (08.09.2016).
- Anthony, Laurence (2014): *AntConc 3.4.0 – Tutorial 7: N-Grams Tool*. Online Publikation: <https://www.youtube.com/watch?v=1FG9N3jqlHY> (08.09.2016).
- Anthony, Laurence (2014): *AntConc 3.4.0 — Tutorial 11: Working with tagged data*. Online Publikation: <https://www.youtube.com/watch?v=gkKna-ka9zw> (25.02.2016).
- Anthony, Laurence (2015): *TagAnt (Version 1.2.0) [Macintosh OS X]*. Tokyo, Japan: Waseda University. Available from <http://www.laurenceanthony.net/> (23.05.2016).
- Anthony, Laurence (2016): *Laurence Anthony's Website – Contact Information*. Online Publikation: <http://www.laurenceanthony.net/contact.html> (07.06.2016).
- Anthony, Laurence (2018): *AntConc (Version 3.5.7) [Macintosh OS X]*. Tokyo, Japan: Waseda University. Online Publikation: <http://www.antlab.sci.waseda.ac.jp/> (14.05.2018).
- Anthony, Laurence (2018): *Spanish Lemma List*. Online Publikation: <http://www.laurenceanthony.net/software/antconc/> (17.05.2018).

6. Bibliographie

- Athelstan (ed.): *MonoConc*. Online Publikation: <http://www.athel.com/mono.html> (21.02.2017).
- Bayerisches Staatsministerium für Bildung und Kultus, Wissenschaft und Kunst (2017): *Für Prüfungszwecke genehmigte Wörterbücher im Bereich der modernen Fremdsprachen*. Online Publikation: https://www.google.de/url?sa=t&rct=j&q=&esrc=s&source=web&cd=3&cad=rja&uact=8&ved=0ahUKEwiOjcuM4PnWAhXjCsAKHQxZAFgQFgg7MAI&url=https%3A%2F%2Fwww.km.bayern.de%2Fdownload%2F11637_liste_genehmigte_wrterbcher_2017_juni.pdf&usg=AOvVaw3enpt99ooGUaerF7LlfZMD (18.10.2017).
- Bayerisches Staatsministerium für Unterricht und Kultus (2007): *Die Seminare in der gymnasialen Oberstufe*. Online Publikation: https://www.isb.bayern.de/download/1581/isb_seminare_komplett_2-aufl.pdf (01.05.2018).
- Blanco Canales, Ana/ Álvarez Martínez, María Ángeles: *Corpus Fono.ele. Acceso al corpus*. Online Publikation: <http://www3.uah.es/fonoele/> (13.10.2017).
- Blanco Canales, Ana/ Álvarez Martínez, María Ángeles: *Corpus Fono.ele. Descripción, objetivos, acceso*. Online Publikation: <http://www3.uah.es/fonoele/corpus.php> (13.10.2017).
- Blanco Canales, Ana/ Álvarez Martínez, María Ángeles: *Corpus Fono.ele. Presentación*. Online Publikation: <http://www3.uah.es/fonoele/index.php> (13.10.2017).
- Boersma, Paul/ Weenink, David: *Praat: doing phonetics by computer*. Online Publikation: <http://www.fon.hum.uva.nl/praat/> (24.11.2016).
- Bosque, Ignacio (2004): *REDES. Diccionario combinatorio del español contemporáneo. Las palabras en su contexto. Presentación del diccionario*. Online Publikation: http://www.rae.es/sites/default/files/PDF_diccionario_redes.pdf (01.10.2017).
- Bougé, Kevin: *Download stop words*. Online Publikation: <https://sites.google.com/site/kevinbouge/stopwords-lists> (12.02.2016).
- Bubenhofen, Noah: *Induktive Statistik: Signifikanztest (2)*. Online Publikation: http://www.bubenhofen.com/korpuslinguistik/kurs/index.php?id=statistik_signifikanzLLR.html (06.09.2016).
- Bubenhofen, Noah: *Statistik: Einführung*. Online Publikation: http://www.bubenhofen.com/korpuslinguistik/kurs/index.php?id=statistik_einfuehrung.html (06.09.2016).
- Bürgel, Christoph: *Forschung*. Online Publikation: <https://kw.uni-paderborn.de/institut-fuer-romanistik/prof-dr-christoph-buergel/forschung/> (02.08.2017).

- Bürgel, Christoph/ Siepmann, Dirk (2015): *Gesamtliste der häufigsten Phraseme des Französischen*. Online Publikation: [https://kw.uni-paderborn.de/fileadmin/fakultaet/Institute/romanistik/Prof. Dr. Christoph Buergel/Gesamtliste_Phraseme_des_Franzoesischen.pdf](https://kw.uni-paderborn.de/fileadmin/fakultaet/Institute/romanistik/Prof._Dr._Christoph_Buergel/Gesamtliste_Phraseme_des_Franzoesischen.pdf) (02.08.2017).
- Bullock, Barbara E./ Toribio, Almeida Jacqueline (2013): *Spanish in Texas Corpus. About the Corpus*. Online Publikation: <http://corpus.spanishintexas.org/en/about-corpus> (06.10.2017).
- Bullock, Barbara E./ Toribio, Almeida Jacqueline (2013): *Spanish in Texas. About the Project*. Online Publikation: <http://spanishintexas.org/about-the-project/> (06.10.2017).
- Bullock, Barbara E./ Toribio, Almeida Jacqueline (2013): *Spanish in Texas Corpus. Interview AF004_1991_EP_SU2011_AD*. Online Publikation: <http://corpus.spanishintexas.org/en/interviews/af0041991epsu2011ad> (06.10.2017).
- Bullock, Barbara E./ Toribio, Almeida Jacqueline (2013): *The Spanish in Texas Corpus Project. COERLL. SpinTX Video Archive. Video 572*. Online Publikation: <http://www.coerll.utexas.edu/spintx/video/572> (06.10.2017).
- Calvo, José (2016): *Textdaten Romanistik. Corpus periodístico Hemero*. Online Publikation: <http://wikis.fu-berlin.de/pages/viewpage.action?pageId=736854944> (14.03.2017).
- CLARIN (ed.): *FCS Aggregator*. Online Publikation: <https://www.clarin.eu/contentsearch> (17.01.2017).
- CLARIN (ed.): *FCS Aggregator. Suche nach ‚presidente‘*. Online Publikation: <https://spraakbanken.gu.se/ws/fcs/2.0/aggregator/#> (09.03.2017).
- C-ORAL-ROM Project (ed.): *Corpus Demo*. Online Publikation: <http://www.elda.org/en/proj/coral/demo/index.html> (22.12.2016).
- C-ORAL-ROM Project (ed.): *Corpus Demo. Diálogo informal, privado (efamdl03)*. Online Publikation: http://www.elda.org/en/proj/coral/demo/spain/Informal/Family_Private/dialogues_conversations/efamdl03.html (10.03.2017).
- Da Silva, Iandra Maria: *La utilización de la lingüística de corpus en la enseñanza de español: usos de pero y sino*. Online Publikation: http://cvc.cervantes.es/ensenanza/biblioteca_ele/publicaciones_centros/PDF/rio_2_006/26_dasilva.pdf (06.06.2017).
- Davies, Mark: *Corpus del Español. Ayuda: Información técnica y créditos*. Online Publikation: <http://www.corpusdelespanol.org/> (28.05.2011).

6. Bibliographie

- Davies, Mark: *Corpus del Español. Ayuda: Sinónimos*. Online Publikation: <http://www.corpusdelespanol.org/> (22.05.2011).
- Davies, Mark: *Corpus del Español. Ayuda: Textos en el corpus*. Online Publikation: <http://www.corpusdelespanol.org/x.asp> (28.05.2011).
- Davies, Mark: *Corpus del Español. Corpora*. Online Publikation: <http://www.corpusdelespanol.org/x.asp> (21.11.2016).
- Davies, Mark: *Corpus del Español – Five Minute Tour*. Online Publikation: <http://www.corpusdelespanol.org/> (20.05.2011).
- Davies, Mark: *Corpus del Español – Género/Histórico*. Online Publikation: <http://www.corpusdelespanol.org/hist-gen/> (19.12.2016).
- Davies, Mark: *Corpus del Español – Género/Histórico (Version 2008)*. Online Publikation: <https://www.corpusdelespanol.org/hist-gen/2008/x.asp> (09.11.2017).
- Davies, Mark: *Corpus del Español. Introducción*. Online Publikation: <http://www.corpusdelespanol.org/> (20.05.2011).
- Davies, Mark: *Corpus del Español. Los corpus – Género/Histórico*. Online Publikation: <http://www.corpusdelespanol.org/xs.asp> (20.12.2016).
- Davies, Mark: *Corpus del Español. New interface*. Online Publikation: <http://www.corpusdelespanol.org/interface2016.asp> (16.01.2017).
- Davies, Mark: *Corpus del Español. Sintaxis: Palabra/Frase*. Online Publikation: http://www.corpusdelespanol.org/help/syntax_s.asp (22.05.2011).
- Davies, Mark: *Corpus del Español. Virtual Corpus*. Online Publikation: <http://www.corpusdelespanol.org/web-dial/> (21.11.2016).
- Davies, Mark: *Corpus del Español – Web/Dialectos*. Online Publikation: <http://www.corpusdelespanol.org/web-dial/> (16.12.2016).
- Davies, Mark: *Corpus del Español – Web/Dialectos. Comparación*. Online Publikation: http://www.corpusdelespanol.org/web-dial/help/compare_s.asp (16.01.2017).
- Davies, Mark: *Corpus.byu.edu. Academic license*. Online Publikation: https://corpus.byu.edu/academic_license.asp?s=q (26.01.2018).
- Davies, Mark: *Corpus.byu.edu. Premium account*. Online Publikation: <https://corpus.byu.edu/premium.asp?c=y> (04.03.2018).
- Davies, Mark: *El corpus del español. Corpora*. Online Publikation: <http://www.corpusdelespanol.org/x.asp?c=4> (20.01.2018).

- Davies, Mark: *El corpus del español. Corpora. NOW*. Online Publikation: <http://www.corpusdelespanol.org/xs.asp?c=4> (24.08.2018).
- Davies, Mark: *El corpus del español. Corpora. Web/Dialectos*. Online Publikation: <https://www.corpusdelespanol.org/xs.asp?c=3> (24.08.2018).
- Davies, Mark: *Full-text corpus data from corpus.byu.edu*. Online Publikation: <http://corpus.byu.edu/full-text/spanish.asp> (13.12.2016).
- Davies, Mark: *Full-text corpus data from corpus.byu.edu – Purchase data*. Online Publikation: <http://corpus.byu.edu/full-text/purchase.asp> (13.12.2016).
- Davies, Mark: *Google Books: Spanish*. Online Publikation: <https://googlebooks.byu.edu/x.asp> (20.01.2018).
- Davies, Mark: *New Spanish and Portuguese Corpora*. Online Publikation: <http://corpus.byu.edu/neh2015.asp> (23.01.2017).
- Davies, Mark: *NOW Corpus (News on the Web)*. Online Publikation: <http://corpus.byu.edu/now/> (16.01.2017).
- Davies, Mark: *NOW Corpus: Spanish*. Online Publikation: <https://www.corpusdelespanol.org/now/> (25.08.2018).
- Davies, Mark: *WordAndPhrase*. Online Publikation: <https://www.wordandphrase.info/span/> (20.01.2018).
- Davies, Mark: *WordAndPhrase. Analyze text*. Online Publikation: <https://www.wordandphrase.info/span/x.asp> (20.01.2018).
- Davies, Mark: *WordAndPhrase. Repartir*. Online Publikation: <https://www.wordandphrase.info/span/x3.asp?d=y&w1=repartir&c1=v> (20.01.2018).
- DGA (ed.2014): *TranscriberAG*. Online Publikation: <http://transag.sourceforge.net> (24.11.2016).
- Diario Correo (ed.): *Cusco celebra hoy escenificación de tradicional Inti Raymi*. Online Publikation: <https://diariocorreo.pe/ciudad/cusco-celebra-hoy-escenificacion-de-tradicional-inti-raymi-527502/> (31.10.2016).
- Directorate General Information Society (ed.): *Integrated reference corpora for spoken romance languages: C-ORAL-ROM – Intro*. Online Publikation: <http://lablita.dit.unifi.it/coralrom/intro.html> (03.12.2011).
- Domínguez, Laura/ Mitchell, Rosamond/ Myles, Florence/ Marsden, Emma: *Spanish Learner Language Oral Corpora. Linguistic development in L2 Spanish. Home Page*. Online Publikation: <http://www.splloc.soton.ac.uk/index.html> (09.10.2017).

6. Bibliographie

- Domínguez, Laura/ Mitchell, Rosamond/ Myles, Florence/ Marsden, Emma: *Spanish Learner Language Oral Corpora. Linguistic development in L2 Spanish. Rationale*. Online Publikation: <http://www.splloc.soton.ac.uk/rationale.html> (12.10.2017).
- Domínguez, Laura/ Mitchell, Rosamond/ Myles, Florence/ Marsden, Emma: *Spanish Learner Language Oral Corpora. Linguistic development in L2 Spanish. Search the SPLLOC Corpus*. Online Publikation: <http://www.splloc.soton.ac.uk/search.php> (12.10.2017).
- Domínguez, Laura/ Mitchell, Rosamond/ Myles, Florence/ Marsden, Emma: *Spanish Learner Language Oral Corpora. Linguistic development in L2 Spanish. Transcript Hermanas*. Online Publikation: <http://www.splloc.soton.ac.uk/data/splloc2/transcripts/Hermanas/H57SMP13.cha> (12.10.2017).
- Domínguez, Laura/ Mitchell, Rosamond/ Myles, Florence/ Marsden, Emma: *Spanish Learner Language Oral Corpora. Linguistic development in L2 Spanish. Search Results*. Online Publikation: <http://www.splloc.soton.ac.uk/dosearch.php> (13.10.2017).
- Domínguez, Laura/ Mitchell, Rosamond/ Myles, Florence/ Marsden, Emma: *SPLLOC. Hermanas story. Audiofile H35NTV13N.mp3*. Online Publikation: <http://www.splloc.soton.ac.uk/data/splloc2/mp3/Hermanas/H35NTV13N.mp3> (11.01.2018).
- Domínguez, Laura/ Mitchell, Rosamond/ Myles, Florence/ Marsden, Emma: *SPLLOC. Hermanas story. Transcript H35NTV13N.cha*. <http://www.splloc.soton.ac.uk/data/splloc2/transcripts/Hermanas/H21NTV10N.cha> (11.01.2018).
- Dr. Dresing & Pehl GmbH (ed.2016): *f4 & f5transkript*. Online Publikation: <http://www.audiotranskription.de/f4.htm> (24.11.2016).
- Duden (ed.): *Wortfamilie*. Online Publikation: <https://www.duden.de/rechtschreibung/Wortfamilie> (16.11.2017).
- Duden (ed.): *Wortfeld*. Online Publikation: <https://www.duden.de/rechtschreibung/Wortfeld> (15.11.2017).
- Ebert, Lars (2012-2017): *XWords*. Online Publikation: <https://www.xwords-generator.de/de> (26.10.2017).
- El montonero (ed.): *El Inti Raymi y el Inca Garcilaso de la Vega*. Online Publikation: <http://elmontonero.pe/cultura/el-inti-raymi-y-el-inca-garcilaso-de-la-vega> (31.10.2016).

- El mundo (2009): '*Save The Children*' ve poco realista tratar de erradicar del trabajo infantil'. Online Publikation: <http://www.elmundo.es/elmundo/2009/06/10/solidaridad/1244656044.html> (17.01.2018).
- Enrique-Arias, Andrés/ Pueyo Mena, F. Javier (2008): *Corpus Biblia Medieval. Contenidos*. Online Publikation: <http://www.bibliamedieval.es/index.php/corpus-contenidos> (20.12.2016).
- Enrique-Arias, Andrés/ Pueyo Mena, F. Javier (2008): *Corpus Biblia Medieval. Página Principal*. Online Publikation: <http://www.bibliamedieval.es> (27.10.2016).
- Enrique-Arias, Andrés/ Pueyo Mena, F. Javier (2008): *Corpus Biblia Medieval. Buscar primer capítulo de Génesis*. Online Publikation: <http://corpus.bibliamedieval.es> (09.03.2017).
- Europarat (ed.2001): Gemeinsamer Europäischer Referenzrahmen für Sprachen (GER). Online Publikation: <http://www.europaeischer-referenzrahmen.de> (06.09.2017).
- Forschungsstelle Eugenio Coseriu (ed.2016): *System, Norm und Rede*. Online Publikation: <http://www.coseriu.de> (12.07.2018).
- Golem (ed.): *Linguee: Suchmaschine für Übersetzungen*. Online Publikation: <http://www.golem.de/0904/66396.html> (14.11.2016).
- González Goñi, Silvia: *Diccionario de la jerga juvenil*. Online Publikation: https://issuu.com/euterpeee/docs/diccionario_jerga (25.02.2018).
- Grup de Recerca Interuniversitari en Aplicacions Lingüístiques (ed.): *Corpus SenSem. Búsqueda*. Online Publikation: <http://grial.uab.es/sensem/corpus?idioma=es> (13.03.2017).
- Grup de Recerca Interuniversitari en Aplicacions Lingüístiques (ed.): *Corpus Trilingüe Paralelo GRIAL. Buscar ,economía' en ,manage/330'*. Online Publikation: <http://grial.uab.es/sensem/cpg/search?page=0&type=norm&text=econom%C3%A0Da&idi=es&docs%5B%5D=8> (10.03.2017).
- Grup de Recerca Interuniversitari en Aplicacions Lingüístiques (ed.): *Inicio*. Online Publikation: <http://grial.uab.es/index.php?idioma=es> (28.10.2016).
- Grup de Recerca Interuniversitari en Aplicacions Lingüístiques (ed.): *Proyectos. Corpus Trilingüe Paralelo GRIAL*. Online Publikation: <http://grial.uab.es/fproj.php?id=3&idioma=es> (21.12.2016).
- Grup de Recerca Interuniversitari en Aplicacions Lingüístiques (ed.): *Recursos. Corpus*. Online Publikation: <http://grial.uab.es/recursos.php?id=1&idioma=es> (21.12.2016).

6. Bibliographie

- Grupo de investigación CLiC (ed.): *AnCora*. Online Publikation: <http://clic.ub.edu/corpus/es/ancora> (12.01.2017).
- Grupo de investigación CLiC (ed.): *AnCora. Búsquedas*. Online Publikation: <http://clic.ub.edu/corpus/es/búsquedas> (12.01.2017).
- Grupo SM (ed.): *Las palabras en su contexto. Diccionario combinatorio práctico del español contemporáneo*. Online Publikation: <http://diccionariopractico.grupo-sm.com> (03.10.2017).
- Hamburger Zentrum für Sprachkorpora (ed.): *EXMARaLDA*. Online Publikation: <http://exmaralda.org/de/> (24.11.2016).
- Hoey, Michael: „Remark at TALC 1998“. In: Lüdeling, A. (2004): *VL: Einführung in die Korpuslinguistik*. Online Publikation: https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/lehre/alte_jahrgaenge/ss-2004/vl-einf-korpling/pdf/VLGeschichte.pdf (09.02.2017).
- Institut Universitari de Lingüística Aplicada (ed.): *BwanaNet. Programa de explotación del Corpus Técnico de IULA*. Online Publikation: <https://bwananet.iula.upf.edu> (26.10.2016).
- Institut Universitari de Lingüística Aplicada (ed.): *BwanaNet. Programa de explotación del Corpus Técnico de IULA. Buscar ‚contaminación‘ en ‚Ciencia y tecnología ambiental y energética‘*. Online Publikation: <http://bwananet.iula.upf.edu/cgi-bin/bwananet0/viewSimple.pl> (13.03.2017).
- Instituto Cervantes (ed.): *Corpus de aprendices de español (CAES)*. Online Publikation: http://www.cervantes.es/lengua_y_ensenanza/tecnologia_espanol/caes.htm (15.10.2017).
- Instituto Cervantes: *Corpus de aprendices de español (CAES). Buscar*. Online Publikation: <http://galvan.usc.es/caes/search> (15.10.2017).
- Instituto Cervantes: *Corpus de aprendices de español (CAES). Buscar ejercicio. Contexto*. Online Publikation: <http://galvan.usc.es/caes/tokens/m3by4yaus279ppo3?lemma1=&lemma2=&lemma3=&lemma4=&tag1=&tag2=&tag3=&tag4=&token1=ejercicio&token2=&token3=&token4=> (15.10.2017).
- Instituto Cervantes: *Corpus de aprendices de español (CAES). Buscar ejercicio. Ejemplos*. Online Publikation: http://galvan.usc.es/caes/search?utf8=✓&level=Cualquiera&mother_tongue=Cualquiera&country=Cualquiera&age_from=&age_to=&sex=Cualquiera&distance=&token1=ejercicio&tag1=&lemma1=&token2=&tag2=&lemma2=&token3=&tag3=&lemma3=&token4=&tag4=&lemma4=&result_type=Ejemplos&ordering=Elemento&page_size=50&commit=Buscar (15.10.2017).

- Instituto Cervantes: *Corpus de aprendices de español (CAES). Buscar ejercicio. Estadística.* Online Publikation: http://galvan.usc.es/caes/search?utf8=✓&level=Cualquiera&mother_tongue=Cualquiera&country=Cualquiera&age_from=&age_to=&sex=Cualquiera&distance=&token1=ejercicio&tag1=&lemma1=&token2=&tag2=&lemma2=&token3=&tag3=&lemma3=&token4=&tag4=&lemma4=&result_type=Estad%C3%ADstica+completa&commit=Buscar (15.10.2017).
- Instituto Cervantes (ed.1997-2018): *Plan curricular del Instituto Cervantes. Niveles de referencia para el español.* Online Publikation: http://cvc.cervantes.es/ensenanza/biblioteca_ele/plan_curricular/ (21.09.2018).
- jEdit (ed.): *jEdit – Programmer's Text Editor.* Online Publikation: <http://jedit.org/index.php?page=download> (07.06.2016).
- John Benjamins Publishing Company (ed.): *C-ORAL-ROM. Integrated Reference Corpora for Spoken Romance Languages.* Online Publikation: <https://www.benjamins.com/#catalog/books/scl.15/main> (22.12.2016).
- Jørgensen, Annette Myre: *El corpus COLA.* Online Publikation: http://www.colam.org/om_prosj-espanol.html (24.02.2018).
- Jørgensen, Annette Myre: *El corpus COLA. Corpus.* Online Publikation: <http://www.colam.org/transkripsjoner-espanol.html> (28.02.2018).
- Jørgensen, Annette Myre: *El corpus COLA. MABPE2-06.* Online Publikation: <http://clu.uni.no/humfak/cola/txt-lyd/mabpe2-06.htm> (17.04.2018).
- Jørgensen, Annette Myre: *El corpus COLA. MABPE2-11b.* Online Publikation: <http://clu.uni.no/humfak/cola/txt-lyd/mabpe2-11b.htm> (17.04.2018).
- Jørgensen, Annette Myre: *El corpus COLA. MAESB2-01b.* Online Publikation: <http://clu.uni.no/humfak/cola/txt-lyd/maesb2-01b.htm> (17.04.2018).
- Jørgensen, Annette Myre: *El corpus COLA. Startseite.* Online Publikation: <http://clu.uni.no/humfak/cola/> (24.02.2018).
- Jørgensen, Annette Myre: *El corpus COLA. Suchmaske.* Online Publikation: <http://clu.uni.no/humfak/cola/s9.html> (24.02.2018).
- Junta de Andalucía (ed.): *Conocer la Alhambra. Generalife.* Online Publikation: <http://www.alhambra-patronato.es/index.php/Generalife/31/0/> (08.12.2016).
- Kohn, Kurt: *Backbone.* Online Publikation: <http://webapps.ael.uni-tuebingen.de/backbone-search/faces/search.jsp> (15.02.2018).
- Kohn, Kurt: *Backbone Project.* Online Publikation: <http://projects.ael.uni-tuebingen.de/backbone/moodle/> (15.02.2018).

6. Bibliographie

- Kohn, Kurt/ Hoffstaedter, Petra (2008): *SACODEYL Corpus-Based Language Learning – Guidelines for Teachers and Learners*. Online Publikation: http://www.um.es/sacodeyl/data/publications/SACODEYL_guidelines_for_teachers.pdf (10.01.2017).
- Koike, Dale: *Spanish Corpus Proficiency Level Training. Intro*. Online Publikation: <http://www.laits.utexas.edu/spt/intro> (05.10.2017).
- Koike, Dale: *Spanish Corpus Proficiency Level Training. Video Activity*. Online Publikation: <http://www.laits.utexas.edu/spt/video> (05.10.2017).
- Laboratorio de Lingüística Informática de la Universidad Autónoma de Madrid (ed.): *Corpus Oral de Referencia de la Lengua Española Contemporánea (CORLEC)*. Online Publikation: <http://www.lllf.uam.es/ESP/Corlec.html> (25.10.2016).
- Laboratorio de Lingüística Informática de la Universidad Autónoma de Madrid (ed.): *Corpus Oral de Referencia de la Lengua Española Contemporánea (CORLEC). Ejemplos*. Online Publikation: <http://www.lllf.uam.es/ESP/Info%20Corlec.html> (10.03.2017).
- Laboratorio de Lingüística Informática de la Universidad Autónoma de Madrid (ed.): *Corpus Oral de Referencia de la Lengua Española Contemporánea (CORLEC). Información y descripción del corpus*. Online Publikation: <http://www.lllf.uam.es/ESP/Info%20Corlec.html> (24.11.2016).
- Laboratorio de Lingüística Informática de la Universidad Autónoma de Madrid (ed.): *Spanish Learner Oral Corpus*. Online Publikation: <http://www.lllf.uam.es/ING/SLOC.html> (11.01.2017).
- Laboratorio de Lingüística Italiana (ed.): *Acceso a las sesiones de C-Or-DiAL*. Online Publikation: <http://lablita.it/app/cordial/corpus.php> (23.02.2018).
- Laboratorio de Lingüística Italiana (ed.): *Acceso a las sesiones de C-Or-DiAL. Búsqueda avanzada*. Online Publikation: <http://lablita.dit.unifi.it/app/cordial/search.php> (19.03.2017).
- Laboratorio de Lingüística Italiana (ed.): *Acceso a las sesiones de C-Or-DiAL. Indices de todo el Corpus*. Online Publikation: <http://lablita.dit.unifi.it/app/cordial/corpus.php> (19.03.2017).
- Laboratorio de Lingüística Italiana (ed.): *Corpus Didáctico Anotado Lingüísticamente (C-Or-DiAL)*. Online Publikation: <http://lablita.dit.unifi.it/corpora/cordial/> (19.03.2017).
- Laurillard, Diana (1998): *How Can Learning Technologies Improve Learning?*. Online Publikation:

<http://web.archive.org/web/20070322002729/http://www.law.warwick.ac.uk/ltj/3-2j.html> (21.07.2011).

Lexical Analysis Software Ltd (ed.): *WordSmithTools*. Online Publikation: <http://www.lexically.net/wordsmith/> (21.02.2017).

Lexical Computing (ed.): *Sketch Engine*. Online Publikation: <https://www.sketchengine.eu> (05.05.2018).

Lexical Computing (ed.): *Sketch Engine. Corpora*. Online Publikation: <https://the.sketchengine.co.uk/auth/corpora/> (07.05.2018).

Lexical Computing (ed.): *Sketch Engine. esTenTen*. Online Publikation: <https://www.sketchengine.eu/esTenTen-spanish-corpus/> (07.05.2018).

Lexical Computing (ed.): *Sketch Engine. esTenTen. Thesaurus: respuesta*. Online Publikation: https://the.sketchengine.co.uk/corpus/thes?corpname=preloaded%2Festenten11_fl5&reload=&lemma=respuesta&lpos=-n&maxthesitems=60&minthesscore=0.0&includeheadword=0&clusteritems=0&minsim=0.15 (07.05.2018).

Lexical Computing (ed.): *Sketch Engine. esTenTen. Word sketch: pregunta*. Online Publikation: https://the.sketchengine.co.uk/corpus/wsketch?corpname=preloaded%2Festenten11_fl5&reload=&lemma=pregunta&lpos=-n&usesubcorp=&minfreq=auto&minscore=0.0&maxitems=25&sort_ws_columns=s&show_lemma_coverage=0&show_lcm=0&show_lcm=1&clustercols=0&minsim=0.15&structured=0&structured=1&min_unary_score=0.0&min_mwlink_freq=100&nr_ws_cols=5&bim_lang= (07.05.2018).

Lexical Computing (ed.): *Sketch Engine. Price List*. Online Publikation: <https://www.sketchengine.eu/price-list/> (05.05.2018).

Linguee GmbH (ed.): *Linguee. Wörterbuch Englisch-Deutsch und Suche in einer Milliarde Übersetzungen*. Online Publikation: <http://www.linguee.de> (14.11.2016).

Llisterri, Joaquim: *Corpus de lengua oral en español*. Online Publikation: http://liceu.uab.es/~joaquim/language_resources/spoken_res/Corp_leng_oral_esp.html (07.03.2017).

Lozano, Cristóbal: *CEDEL2 info*. Online Publikation: <http://www.uam.es/proyectosinv/woslac/collaborating.htm> (16.04.2018).

Lozano, Cristóbal: *CEDEL2. Es_wr61_1_mda*. Online Publikation: http://cedel2.learnercorpora.com/student/es_wr_61_1_mda/?search_term=paisaje (16.04.2018).

6. Bibliographie

- Marcos Marín, Francisco/ Faulhaber, Charles B.: *ADMYTE. Contenido*. Online Publikation: <http://www.admyte.com/admyteonline/contenido.htm> (27.10.2016).
- Marcos Marín, Francisco/ Faulhaber, Charles B.: *ADMYTE. Historia*. Online Publikation: <http://www.admyte.com/admyteonline/historia.htm> (27.10.2016).
- Marcos Marín, Francisco/ Faulhaber, Charles B.: *ADMYTE. Presentación*. Online Publikation: <http://www.admyte.com/admyteonline/presentacion.htm> (27.10.2016).
- Marcos Marín, Francisco (1998): *La Colección Foulché-Delbosc de la Biblioteca Nacional de la República Argentina. Catalogación y Edición Electrónica*. Online Publikation: <http://www.illf.uam.es/~fmarcos/informes/BNArgentina/fd01.htm> (06.06.2017).
- Meurer, Paul: *Corpuscle. COLA*. Online Publikation: <http://clarino.uib.no/korpuskel/clarino-metadata?session-id=244347694754873&corpus=cola&default-corpus=cola&resource=cola> (24.02.2018).
- Mitchell, Rosamond/ Richard, Laurence/ Romero de Mills, Patricia: *The LANGSNAP Project: Languages and Social Networks Abroad*. Online Publikation: <http://langsnap.soton.ac.uk/project.html> (12.10.2017).
- Mitchell, Rosamond/ Richard, Laurence/ Romero de Mills, Patricia: *The LANGSNAP Project: Writing (Drugs) – Participant 150. Full Record*. Online Publikation: <http://langsnap.soton.ac.uk/849/> (12.10.2017).
- Mitchell, Rosamond/ Richard, Laurence/ Romero de Mills, Patricia: *The LANGSNAP Project: Writing (Drugs) – Participant 150. Text (Transcript with Grammar Analysis)*. Online Publikation: <http://langsnap.soton.ac.uk/850/2/F150c.mor.pst.cex> (12.10.2017).
- Mitchell, Rosamond/ Richard, Laurence/ Romero de Mills, Patricia: *Welcome to LANGSNAP*. Online Publikation: <http://langsnap.soton.ac.uk> (12.10.2017).
- Molino de Ideas (ed.): *Molino de Ideas. Tecnología lingüística y Procesamiento de Lenguaje Natural*. Online Publikation: <http://molinodeideas.es/index.html> (14.03.2017).
- Molino de Ideas (ed.): *Hemero.es - El corpus molinero*. Online Publikation: <http://www.hemero.es> (14.03.2017).
- Molino de Ideas (ed.): *Hemero.es - El corpus molinero – Buscar ,Una palabra o una frase' – naturaleza*. Online Publikation: [http://www.hemero.es/#1&espania\\$&texto&naturaleza](http://www.hemero.es/#1&espania$&texto&naturaleza) (14.03.2017).

- Molino de Ideas (ed.): *Hemero.es - El corpus molinero – Buscar ,Una palabra y sus palabras parecidas’ – ,muy parecido’ – naturaleza*. Online Publikation: [http://www.hemero.es/#3&espania\\$&texto&naturaleza&0.80](http://www.hemero.es/#3&espania$&texto&naturaleza&0.80) (14.03.2017).
- Molino de Ideas (ed.): *Hemero.es - El corpus molinero – Buscar ,Una palabra y sus palabras parecidas’ – ,muy poco parecido’ – naturaleza*. Online Publikation: [http://www.hemero.es/#3&espania\\$&texto&naturaleza&0.20](http://www.hemero.es/#3&espania$&texto&naturaleza&0.20) (14.03.2017).
- Moreno Fernández, Francisco: *Proyecto para el Estudio Sociolingüístico del Español de España y de América. Consulta Corpus PRESEEA*. Online Publikation: <http://preseea.linguas.net/Corpus.aspx> (27.10.2016).
- Moreno Fernández, Francisco: *Proyecto para el Estudio Sociolingüístico del Español de España y de América. Presentación*. Online Publikation: <http://preseea.linguas.net/Inicio.aspx> (27.10.2016).
- Nerbonne, John: *Natural Language Processing in Computer-Assisted Language Learning*. Online Publikation: <http://www.let.rug.nl/nerbonne/papers/nlp-hndbk-call.pdf> (29.10.2017), S. 5.
- Orihuela, José Luis (2006): *Diccionario combinatorio práctico del español contemporáneo*. Online Publikation: <https://www.ecuaderno.com/2006/12/07/diccionario-combinatorio-practico-del-espanol-contemporaneo/> (13.10.2017).
- Pascher, Franziska (2015): *Spanische Korpora und Textdatenbanken*. Online Publikation: <http://wikis.fu-berlin.de/display/tdr/Spanische+Korpora+und+Textdatenbanken> (06.03.2017).
- Pérez-Paredes, Pascual: *Sacodeyl. Category Tree*. Online Publikation: <http://sacodeyl.inf.um.es/sacodeyl-search2/faces/search.jsp> (06.09.2017).
- Pérez-Paredes, Pascual: *Sacodeyl. European Youth Language*. Online Publikation: <http://sacodeyl.inf.um.es/sacodeyl-search2/> (18.11.2016).
- Pérez-Paredes, Pascual: *Sacodeyl. European Youth Language. Search ,gustaría’*. Online Publikation: <http://sacodeyl.inf.um.es/sacodeyl-search2/faces/search.jsp> (11.03.2017).
- Pérez-Paredes, Pascual: *Sacodeyl. Project Overview*. Online Publikation: <http://www.um.es/sacodeyl/> (24.01.2017).
- Pons (ed.): *Online Wörterbuch Deutsch – Spanisch*. Online Publikation: <http://de.pons.com> (03.11.2017).
- Pontificia Universidad Católica de Valparaíso (ed.): *El Grial*. Online Publikation: <http://www.elgrial.cl/index2015.php> (26.10.2016).

6. Bibliographie

- Pontificia Universidad Católica de Valparaíso (ed.): *El Grial. El Manchador de Textos*. Online Publikation: <http://elgrial.cl/grial/#el-manchador> (06.03.2017).
- Pontificia Universidad Católica de Valparaíso (ed.): *El Grial. Interfaz de etiquetaje e interrogación de corpus textuales*. Online Publikation: <http://www.elgrial.cl> (26.10.2016).
- PRESEEA (ed.2014-2017): *Corpus del Proyecto para el estudio sociolingüístico del español de España y de América*. Alcalá de Henares: Universidad de Alcalá. Online Publikation: <http://preseea.linguas.net> (09.01.2017).
- PRESEEA (ed.2014-2017): *Corpus del Proyecto para el estudio sociolingüístico del español de España y de América. Corpus*. Alcalá de Henares: Universidad de Alcalá. Online Publikation: <http://preseea.linguas.net/Corpus.aspx> (14.03.2017).
- Pusch, Claus D. (2014): *A survey of spoken language corpora in Romance*. Online Publikation: http://www.romanistik.uni-freiburg.de/pusch/Download/publikationen/2002_Survey_corpora.pdf (07.03.2017).
- Real Academia Española (ed.): *Banco de datos. CDH*. Online Publikation: <http://www.rae.es/recursos/banco-de-datos/cdh> (13.01.2017).
- Real Academia Española (ed.): *Banco de datos. CORDE*. Online Publikation: <http://www.rae.es/recursos/banco-de-datos/corde> (23.01.2017).
- Real Academia Española (ed.): *Banco de datos. CORDE. Información*. Online Publikation: <http://www.rae.es/rae/gestores/gespub000019.nsf/voTodosporId/B4E26FC2520104D8C125716400455C06?OpenDocument&i=1> (19.05.2011).
- Real Academia Española (ed.): *Banco de datos. CORDE. Tamaño y criterios de selección*. Online Publikation: [http://www.rae.es/rae/gestores/gespub000019.nsf/\(voAnexos\)/arch475E744872738671C125716500381CF8/\\$FILE/TamanoycriteriosCORDE.htm](http://www.rae.es/rae/gestores/gespub000019.nsf/(voAnexos)/arch475E744872738671C125716500381CF8/$FILE/TamanoycriteriosCORDE.htm) (19.05.2011).
- Real Academia Española (ed.): *Banco de datos. CORPES*. Online Publikation: <http://www.rae.es/recursos/banco-de-datos/corpes-xxi> (24.10.2016).
- Real Academia Española (ed.): *Banco de datos. CREA*. Online Publikation: <http://www.rae.es/recursos/banco-de-datos/crea> (12.12.2016).
- Real Academia Española (ed.): *Banco de datos. CREA Escrito*. Online Publikation: <http://www.rae.es/recursos/banco-de-datos/crea-escrito> (14.12.2016).
- Real Academia Española (ed.): *Banco de datos. CREA Oral*. Online Publikation: <http://www.rae.es/rae/gestores/gespub000019.nsf/voTodosporId/F651954693D3AC5EC125716400426004?OpenDocument> (20.05.2011).

- Real Academia Española (ed.): *Banco de datos. Nueva version del CREA (textos de junio 2008)*. Online Publikation: <http://www.rae.es/rae/gestores/gespub000019.nsf/voTodosporId/E1E307BE0FBE E6FFC1257464003B0B63?OpenDocument> (20.05.2011).
- Real Academia Española (ed.): *CDH*. Online Publikation: <http://web.frl.es/CNDHE/view/inicioExterno.view> (13.01.2017).
- Real Academia Española (ed.): *CDH. Entrada Compleja*. Online Publikation: <http://web.frl.es/CNDHE/org/publico/pages/consulta/entradaCompleja.view> (17.04.2018).
- Real Academia Española (ed.): *CORDE*. Online Publikation: <http://corpus.rae.es/cordenet.html> (09.03.2017).
- Real Academia Española (ed.): *CORDE. Estadística de 'televisión'*. Online Publikation: <http://corpus.rae.es/cgi-bin/crpsrvEx.dll> (09.03.2017).
- Real Academia Española (ed.): *CORPES*. Online Publikation: <http://web.frl.es/CORPES/view/inicioExterno.view> (24.10.2016).
- Real Academia Española (ed.): *CORPES. Buscar 'libros'*. Online Publikation: <http://web.frl.es/CORPES/org/publico/pages/consulta/entradaCompleja.view> (11.03.2017).
- Real Academia Española (ed.): *CORPES. Coapariciones*. Online Publikation: <http://web.frl.es/CORPES/org/publico/pages/colocacion/coaparicion.view> (09.11.2017).
- Real Academia Española (ed.): *CORPES. Descripción del sistema de codificación*. Online Publikation: http://www.rae.es/sites/default/files/CORPES_Sistema_de_codificacion_12_2015.pdf (24.10.2016).
- Real Academia Española (ed.): *CORPES. Entrada Compleja*. Online Publikation: <http://web.frl.es/CORPES/org/publico/pages/consulta/entradaCompleja.view> (17.04.2018).
- Real Academia Española (ed.): *CORPES. Manual de uso*. Online Publikation: <http://web.frl.es/CORPES/org/publico/pages/ayuda/manual.view> (20.10.2017), S. 17-18.
- Real Academia Española (ed.): *CREA (Version 2015)*. Online Publikation: <http://corpus.rae.es/creanet.html> (19.05.2011).
- Real Academia Española (ed.): *CREA. Ayuda*. Online Publikation: http://corpus.rae.es/ayuda_c.htm (14.05.2011).

6. Bibliographie

- Real Academia Española (ed.): *CREA. Buscar ,gata'.* Online Publikation: <http://corpus.rae.es/cgi-bin/crpsrvEx.dll> (24.02.2017).
- Real Academia Española (ed.): *CREA Escrito y Oral.* Online Publikation: <http://www.rae.es/rae/gestores/gespub000019.nsf/voTodosporId/B104F9F0D0029604C1257164004032BE?OpenDocument&i=1> (19.05.2011).
- Real Academia Española (ed.): *CREA. Estadísticas.* Online Publikation: <http://buscon.rae.es/nomina/SrvltGUIBusTextos?est=1> (14.05.2011).
- Real Academia Española (ed.): *CREA – Listado de frecuencias.* Online Publikation: <http://corpus.rae.es/lfrecuencias.html> (22.07.2015).
- Real Academia Española (ed.): *CREA. Listado de frecuencias. 10000 formas más frecuentes.* Online Publikation: http://corpus.rae.es/frec/10000_formas.TXT (24.02.2017).
- Real Academia Española (ed.): *CREA. Versión anotada.* Online Publikation: <http://web.frl.es/CREA/view/inicioExterno.view> (17.04.2018).
- Real Academia Española (ed.): *CREA. Versión anotada. Entrada Compleja.* Online Publikation: <http://web.frl.es/CREA/org/publico/pages/consulta/entradaCompleja.view> (17.04.2018).
- Real Academia Española (ed.): *La RAE y Banco Santander renuevan su convenio para el CORPES XXI.* Online Publikation: <http://www.rae.es/noticias/la-rae-y-banco-santander-renuevan-su-convenio-para-el-corpes-xxi> (14.12.2016).
- Research and Development Unit for English Studies, Birmingham City University (ed.): *WebCorp Linguist's Search Engine.* Online Publikation: <http://wse1.webcorp.org.uk/home/guide.html> (02.09.2018).
- Research and Development Unit for English Studies, Birmingham City University (ed.): *WebCorp Live. Concordance the web in real-time.* Online Publikation: <http://www.webcorp.org.uk/live/index.jsp> (07.12.2016).
- Research and Development Unit for English Studies, Birmingham City University (ed.): *WebCorp Live. Concordance the web in real-time. Guide.* Online Publikation: <http://www.webcorp.org.uk/live/guide.jsp> (08.12.2016).
- Research and Development Unit for English Studies, Birmingham City University (ed.): *WebCorp Live. Concordance the web in real-time. Guide – How does it work?.* Online Publikation: <http://www.webcorp.org.uk/live/guide-how.jsp> (08.12.2016).

Research and Development Unit for English Studies, Birmingham City University (ed.):
WebCorp Live. Concordance the web in real-time. Guide - Post. Online
Publikation: <http://www.webcorp.org.uk/live/guide-post.jsp> (08.12.2016).

Research and Development Unit for English Studies, Birmingham City University (ed.):
WebCorp Live. Concordance the web in real-time. Live. Online Publikation:
<http://www.webcorp.org.uk/live/> (29.04.2017).

Research and Development Unit for English Studies, Birmingham City University (ed.):
WebCorp Live. Concordance the web in real-time. Live. Index. Online Publikation:
<http://www.webcorp.org.uk/live/index.jsp> (08.12.2016).

Research and Development Unit for English Studies, Birmingham City University (ed.):
WebCorp Live. Concordance the web in real-time. Suche nach ‚app‘ mittels der Suchmaschine Bing. Online Publikation:
<http://www.webcorp.org.uk/live/cache.jsp?id=FNyVWM62db&info=on&span=c50&colls=on&sort=none&sortpos=1&case=on&date=x> (08.12.2016).

Research and Development Unit for English Studies, Birmingham City University (ed.):
WebCorp Live. Concordance the web in real-time. Wordlist einer Internetseite zum Generalife mit N-Gramm-Größe 1. Online Publikation:
<http://www.webcorp.org.uk/live/wdlist.jsp?url=http%3A%2F%2Fwww.alhambra-patronato.es%2Findex.php%2FGeneralife%2F31%2F0%2F&sw=on&i=on&n=1>
(08.12.2016).

Research and Development Unit for English Studies, Birmingham City University (ed.):
WebCorp Live. Concordance the web in real-time. Wordlist einer Internetseite zum Generalife mit N-Gramm-Größe 5. Online Publikation:
<http://www.webcorp.org.uk/live/wdlist.jsp?url=http%3A%2F%2Fwww.alhambra-patronato.es%2Findex.php%2FGeneralife%2F31%2F0%2F&sw=on&i=on&n=1>
(08.12.2016).

Research and Development Unit for English Studies, Birmingham City University (ed.):
WebCorp Live. Concordance the web in real-time. Wordlist Tool. Online
Publikation: <http://www.webcorp.org.uk/live/wdlist.jsp> (08.12.2016).

Rieger, Burghard (1979): *Repräsentativität: von der Unangemessenheit eines Begriffs zur Kennzeichnung eines Problems linguistischer Korpusbildung.* Online Publikation:
<https://www.uni-trier.de/fileadmin/fb2/LDV/Rieger/Publikationen/Aufsaeetze/79/rub79.pdf>
(19.05.2017).

Sánchez, Fernando/ Porta, Jordi/ Sancho, José Luis/ Nieto, Amalio/ Ballester, Almudena/ Fernández, Adelaida/ Gómez, Javier/ Gómez, Laura/ Raigal, Encarnación/ Ruiz, Rafael: *La anotación de los corpus CREA y CORDE.* Online Publikation:
<http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/3485/1973>
(21.12.2016).

6. Bibliographie

- Schmid, Helmut: *TreeTagger – Spanish Tagset*. Online Publikation: <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/spanish-tagset.txt> (24.05.2016).
- Schmidt, Thomas/ Wörner, Kai/ Lehmberg, Timm/ Hedeland, Hanna: *EXACT*. Online Publikation: <http://exmaralda.org/de/exakt-de/> (21.02.2017).
- Scott, Mike/ Johns, Tim: *Free Software. Microconcord*. Online Publikation: <http://lexically.net/software/index.htm> (15.07.2017).
- Seminario de Lingüística Informática (SLI) da Universidade de Vigo (ed.): *Corpus CLUVI – Información sobre o Corpus CLUVI*. Online Publikation: <http://sli.uvigo.es/CLUVI/index.php> (20.08.2018).
- Seminario de Lingüística Informática (SLI) da Universidade de Vigo (ed.): *CLUVI Parallel Corpus – Buscar ,pescado' en el corpus CONSUMER*. Online Publikation: <http://sli.uvigo.gal/CLUVI/index.php?corpus=20&tipo=14&lang=en> (06.03.2018).
- Seminario de Lingüística Informática (SLI) da Universidade de Vigo (ed.): *CLUVI Parallel Corpus – More Información. Introduction*. Online Publikation: http://sli.uvigo.es/CLUVI/info_en.html (23.01.2017).
- Seminario de Lingüística Informática (SLI) da Universidade de Vigo (ed.): *Corpus Lingüístico da Universidade de Vigo: Pescudas no Corpus Paralelo CLUVI*. Online Publikation: <http://sli.uvigo.es/CLUVI/> (03.08.2011).
- Seminario de Lingüística Informática (SLI) da Universidade de Vigo (ed.): *Search the CLUVI Parallel Corpus. Corpus CLUVI sections*. Online Publikation: http://sli.uvigo.es/CLUVI/index_en.html (23.01.2017).
- Simov, Kiril/ Osenova, Petya: *CLaRK – an XML Based System for Corpora Development*. Online Publikation: <http://www.bultreebank.org/clark/index.html> (24.11.2016).
- Sourceforge (ed.): *Xaira*. Online Publikation: <http://xaira.sourceforge.net> (21.02.2017).
- Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Das Gymnasium in Bayern*. Online Publikation: <http://www.isb-gym8-lehrplan.de/contentserv/3.1.neu/g8.de/index.php?StoryID=26350> (07.09.2017).
- Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Fachprofil Moderne Fremdsprachen*. Online Publikation: <http://www.isb-gym8-lehrplan.de/contentserv/3.1.neu/g8.de/index.php?StoryID=26366> (08.09.2017).
- Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Fachprofil Spanisch*. Online Publikation: <http://www.isb-gym8-lehrplan.de/contentserv/3.1.neu/g8.de/index.php?StoryID=26376> (08.09.2017).

- Staatsinstitut für Schulqualität und Bildungsforschung München (ISB): *Jahrgangsstufenarbeiten (Gymnasium). Englisch*. Online Publikation: <https://www.isb.bayern.de/gymnasium/leistungserhebungen/jahrgangsstufenarbeiten-eng-gymnasium/englisch/> (09.01.2018).
- Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Spanisch (3. Fremdsprache) – Jahrgangsstufe 8*. Online Publikation: <http://www.isb-gym8-lehrplan.de/contentserv/3.1.neu/g8.de/index.php?StoryID=26253> (06.09.2017).
- Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Spanisch (3. Fremdsprache) – Jahrgangsstufe 9*. Online Publikation: <http://www.isb-gym8-lehrplan.de/contentserv/3.1.neu/g8.de/index.php?StoryID=26220> (06.09.2017).
- Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Spanisch (3. Fremdsprache) – Jahrgangsstufe 10*. Online Publikation: <http://www.isb-gym8-lehrplan.de/contentserv/3.1.neu/g8.de/index.php?StoryID=26431> (06.09.2017).
- Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Spanisch (3. Fremdsprache) – Jahrgangsstufe 11/12*. Online Publikation: <http://www.isb-gym8-lehrplan.de/contentserv/3.1.neu/g8.de/index.php?StoryID=26514> (06.09.2017).
- Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Spanisch (spätbeginnend) – Jahrgangsstufe 10*. Online Publikation: <http://www.isb-gym8-lehrplan.de/contentserv/3.1.neu/g8.de/index.php?StoryID=26804> (06.09.2017).
- Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Spanisch (spätbeginnend) – Jahrgangsstufe 11*. Online Publikation: <http://www.isb-gym8-lehrplan.de/contentserv/3.1.neu/g8.de/index.php?StoryID=26805> (06.09.2017).
- Staatsinstitut für Schulqualität und Bildungsforschung München (ISB) (ed.2004): *Spanisch (spätbeginnend) – Jahrgangsstufe 12*. Online Publikation: <http://www.isb-gym8-lehrplan.de/contentserv/3.1.neu/g8.de/index.php?StoryID=26806> (06.09.2017).
- Subirats, Carlos/ Ortega, Marc: *Corpus del Español Actual (CEA)*. Online Publikation: <http://spanishfn.org/tools/cea/spanish> (17.03.2017).
- Subirats, Carlos/ Ortega, Marc: *Corpus del Español Actual (CEA). Word lookup – Buscar ,trabajo'*. Online Publikation: <http://sfncorpora.uab.es/CQPweb/cea/redirect.php?redirect=concBreakdownBoth&qname=f8b81grfyp&pp=50&uT=y> (17.03.2017).

6. Bibliographie

- TALN, DiCE (ed.): *HARenEs. ¡Aprende colocaciones!*. Online Publikation: <http://harenes.taln.upf.edu/CakeHARenEs/> (14.10.2017).
- Text Encoding Initiative Consortium (ed.2016): *TEI: Text Encoding Initiative*. Online Publikation: <http://www.tei-c.org/index.xml> (24.11.2016).
- The Linguist List (ed.): *Texts & Corpora related to Spanish*. Online Publikation: <http://linguistlist.org/sp/SearchWRListing-action.cfm?subclassid=5727&SearchType=SL&WRTypeID=1> (07.03.2017).
- tsmorton (ed.2004): *WordFreak*. Online Publikation: <http://wordfreak.sourceforge.net> (24.11.2016).
- Uni Bremen Blogs (ed.): *Linguistische Werkstatt. Spanische Korpora*. Online Publikation: <https://blogs.uni-bremen.de/lingwerk/spanische-korpora/> (06.03.2017).
- Uniovi (ed.): *Spanish Corpus Search*. Online Publikation: <http://www.cs.upc.edu/~nlp/tools/corpus-es.php> (27.01.2017).
- Universidad Autónoma de Madrid (ed.): *Corpus de Español como Lengua Extranjera (ELE)*. Online Publikation: <http://cartago.llf.uam.es/corele/index.html> (12.01.2017).
- Universidad Autónoma de Madrid (ed.): *Corpus de Español como Lengua Extranjera (ELE). Resultados de errores de verbo en estudiantes cuya lengua materna es el alemán*. Online Publikation: http://cartago.llf.uam.es/exist/rest/db/corpus/L1_y_cat.xql?category=V&L1=German (10.03.2017).
- Universidad de Santiago de Compostela (ed.2001): *Arthus. Buscar verbos con ,hasta'*. Online Publikation: <http://www.bds.usc.es/consultas/preposicion.html> (13.03.2017).
- Universidad de Santiago de Compostela (ed.2001): *Arthus. Composición del corpus*. Online Publikation: <http://www.bds.usc.es/corpus.html> (23.11.2016).
- Universidad de Santiago de Compostela (ed.2001): *Base de datos sintácticos del español actual*. Online Publikation: <http://www.bds.usc.es> (25.10.2016).
- Universidade de Vigo. Grupo de Investigación Talg (ed.): *CLUVI Parallel Corpus*. Online Publikation: http://lod.iula.upf.edu/resources/metadata_CorpusCLUVIV2 (22.11.2016).
- Universität Leipzig - Institut für Informatik - Abteilung ASV (1998-2017): *Wortschatz Universität Leipzig. Spa_newscrawl_2011*. Online Publikation: http://corpora.uni-leipzig.de/de?corpusId=spa_newscrawl_2011 (21.10.2017).

Universität Leipzig - Institut für Informatik - Abteilung ASV (1998-2017): *Wortschatz Universität Leipzig. Spa_newscrawl_2011. Suche nach ,cielo'*. Online Publikation: http://corpora.uni-leipzig.de/de/res?corpusId=spa_newscrawl_2011&word=cielo (10.11.2017).

Universität Leipzig - Institut für Informatik - Abteilung ASV (1998-2017): *Wortschatz Universität Leipzig. Spa_newscrawl_2011. Suche nach ,garrafal'*. Online Publikation: http://corpora.uni-leipzig.de/de/res?word=garrafal&corpusId=spa_newscrawl_2011 (21.10.2017).

Universitat de Barcelona (ed.2016): *LEXESP. Léxico informatizado del español*. Online Publikation: <http://www.publicacions.ub.edu/ficha.aspx?cod=05096e> (22.12.2016).

Vega, Juan José/ Guzmán Palomino, Luis (2005): *El Inti Raymi Inkaico. La verdadera historia de la gran fiesta del sol*. Online Publikation: http://sisbib.unmsm.edu.pe/Bibvirtual/Publicaciones/antropologia/2003_N01/a05.htm# (31.10.2016).

VISL (ed.): *Corpus Search Interface*. Online Publikation: https://beta.visl.sdu.dk/corpus_linguistics.html#Spanish_corpora (05.11.2016).

VISL (ed.): *Europarl-es*. Online Publikation: <http://corp.hum.sdu.dk/cqp.es.html> (05.12.2016).

VISL (ed.): *Europarl-es. Buscar ,ojo'*. Online Publikation: http://corp.hum.sdu.dk/cgi-bin/cqp.cgi?utf8=✓&lang=es&corpus=HIS_EUROPARL&search=ojo&type=basic_search&case=insensitive&Action=Search&password= (13.03.2017).

VISL (ed.): *Guided Tour*. Online Publikation: <https://beta.visl.sdu.dk/guide/> (05.12.2016).

VISL (ed.): *Internet with semantic roles*. Online Publikation: <http://corp.hum.sdu.dk/cqp.es.html> (05.12.2016).

VISL (ed.): *Internet with semantic roles. Buscar ,profesor' & ,agent'*. Online Publikation: http://corp.hum.sdu.dk/cgi-bin/cqp.cgi?corpus=HIS_WEB_ROLE&lang=es&utf8=✓&type=refine_search&search=%5Bword%3D%22profesor%22+%26+role%3D%22%28.*+%29%3FAG.*%28+.*%29%3F%22+%25c%5D (13.03.2017).

VISL (ed.): *Internet with semantic roles. Buscar ,profesor' & ,patient'*. Online Publikation: http://corp.hum.sdu.dk/cgi-bin/cqp.cgi?corpus=HIS_WEB_ROLE&lang=es&utf8=✓&type=refine_search&search=%5Bword%3D%22profesor%22+%26+role%3D%22%28.*+%29%3FPAT%28+.*%29%3F%22+%25c%5D (13.03.2017).

6. Bibliographie

VISL (ed.): *Internet with semantic roles. Buscar ‚profesor‘ & ‚speaker‘*. Online Publikation: http://corp.hum.sdu.dk/cgi-bin/cqp.cgi?corpus=HIS_WEB_ROLE&lang=es&utf8=✓&type=refine_search&search=%5Bword%3D%22profesor%22+%26+role%3D%22%28.*+%29%3FSP%28+.*%29%3F%22+%25c%5D (13.03.2017).

Weblicht (ed.) (2016): *FCS Aggregator*. Online Publikation: http://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/FCS_Aggregator (17.01.2017).

Wicher, Oliver: *Projekte. Schulfranzösisch – das Lehrbuchkorpus FFLTC*. Online Publikation: <https://kw.uni-paderborn.de/institut-fuer-romanistik/oliver-wicher/projekte/> (17.08.2017).

Wikilengua del español (ed.): *Jerga juvenil / España*. Online Publikation: http://www.wikilengua.org/index.php/Jerga_juvenil/España (25.02.2018).

6.4. Schriftlicher und telefonischer Kontakt

E-Mail-Kontakt mit dem Ersteller-Team des Korpus *Sacodeyl* am 14.02.2018.

E-Mail-Kontakt mit der Redaktion ‚Französisch‘ des Cornelsen Verlags am 28.07.2017.

E-Mail-Kontakt mit Juan Manuel García von *Micronet* am 14.03.2017.

E-Mail-Kontakt mit Heike Malinowski, Redaktionsleiterin ‚Weitere Fremdsprachen‘ des Cornelsen Verlags, am 31.07.2017.

Telefonat mit Heike Malinowski, Redaktionsleiterin ‚Weitere Fremdsprachen‘ des Cornelsen Verlags, am 16.08.2017.

Telefonat mit Heike Malinowski, Redaktionsleiterin ‚Weitere Fremdsprachen‘ des Cornelsen Verlags, am 23.08.2017.

7. Abbildungsverzeichnis

Abbildung 1: System – Norm – Rede nach Eroms (2000).....	19
Abbildung 2: Beginn der Liste der 10.000 häufigsten Lexeme des CREA.....	59
Abbildung 3: Beispiel für Konkordanzen des Lexems <i>gata</i> im CREA.....	60
Abbildung 4: Übersicht über die gängigsten regulären Ausdrücke.....	62
Abbildung 5: Ergebnisse der Stichwortsuche in den Korpora CdE:Nuevo, CREA und CORPES (absolut).	67
Abbildung 6: Ergebnisse der Stichwortsuche in den Korpora CdE:Nuevo, CREA und CORPES (normalisiert je eine Mio. Lexeme).	68
Abbildung 7: CREA – Suchmaske der annotierten Version des CREA bei der Suche nach der Präposition <i>para</i>	78
Abbildung 8: CREA – Anzeige der Konkordanzen für <i>para</i> als Präposition in der annotierten Version des CREA.	79
Abbildung 9: CORPES – Suche nach dem Substantiv <i>ordenador</i> im mündlichen Subkorpus.	81
Abbildung 10: CORPES – Anzeige des Suchbegriffs <i>ordenador</i> im jeweiligen sprachlichen Kontext inklusive Audiodatei.	82
Abbildung 11: CdE:Orig – Suche nach Substantiven, die zusammen mit den Verben <i>empezar</i> oder <i>comenzar</i> erscheinen.	84
Abbildung 12: CdE:Orig – Ergebnis der Suche nach Substantiven, die zusammen mit <i>empezar</i> oder <i>comenzar</i> verwendet werden.	85
Abbildung 13: COLA – Ausschnitt aus dem Transkript eines Gesprächs zwischen drei Jugendlichen.	88
Abbildung 14: <i>Spanish in Texas Corpus</i> – Interview mit einer 20-jährigen Frau aus El Paso.	90
Abbildung 15: <i>SpinTX Video Archive</i> – Transkript eines Interviewausschnitts aus dem <i>Spanish in Texas Corpus</i>	91
Abbildung 16: <i>Spa_news_2011_3M</i> – Präsentation der Ergebnisse zum Suchbegriff <i>presidente</i>	93
Abbildung 17: <i>Spa_newscrawl_2011</i> – Ergebnis der Suche nach dem Adjektiv <i>garrafal</i>	94
Abbildung 18: <i>Spa_newscrawl_2011</i> – Graph zum Adjektiv <i>garrafal</i>	94
Abbildung 19: CORDE – Suche nach <i>televisión</i> im gesamten Korpus.....	99
Abbildung 20: CORDE – Statistik der Ergebnisse des Substantivs <i>televisión</i>	99
Abbildung 21: CDH – Suche nach dem Lemma <i>idioma</i>	100
Abbildung 22: CDH – Ergebnis der Suche nach Kollokationen mit dem Substantiv <i>idioma</i>	101
Abbildung 23: CdE:Orig (Gesamt) – Benutzeroberfläche bei der Suche nach dem Substantiv <i>teléfono</i>	103

7. Abbildungsverzeichnis

Abbildung 24: CdE:Orig (Gesamt) – Übersicht über das Erscheinen des Substantivs <i>teléfono</i> .	103
Abbildung 25: <i>Biblia Medieval</i> – Parallele Darstellung einiger Schriften zum ersten Kapitel des Buchs Genesis – Abschnitt I.	104
Abbildung 26: <i>Biblia Medieval</i> – Parallele Darstellung einiger Schriften zum ersten Kapitel des Buchs Genesis – Abschnitt II.	105
Abbildung 27: CORELE – Analyse von Fehlern deutscher Muttersprachler im Hinblick auf Verbformen.	109
Abbildung 28: SPLLOC – Transkript einer Geschichte zur Aufgabe <i>Dos Hermanas</i> .	111
Abbildung 29: <i>Fono.ele</i> – Konversationen von deutschen Spanischlernern.	112
Abbildung 30: LANGSNAP – Transkript eines argumentativen Aufsatzes zum Thema <i>Drugs</i> inklusive Grammatikanalyse.	114
Abbildung 31: CAES – Konkordanzen des Substantivs <i>ejercicio</i> .	115
Abbildung 32: CAES – Substantiv <i>ejercicio</i> innerhalb des Texts einer Portugiesin mit dem Sprachniveau A2.	115
Abbildung 33: CEDEL2 – Aufsatz einer 61-jährigen Spanierin über ihre Heimat.	117
Abbildung 34: <i>Corpus Trilingüe Paralelo Grial</i> – Ergebnisse der Suche nach <i>economía</i> in einem Dokument über Management-Informationssysteme.	121
Abbildung 35: CLUVI – Ergebnis des CONSUMER-Korpus bei der Suche nach <i>pescado</i> .	123
Abbildung 36: CEA – Ergebnisse der Suche nach <i>trabajo</i> in der Funktion <i>Word Lookup</i> .	126
Abbildung 37: CEA – Häufigkeit der Ergebnisse der Suche nach <i>trabajo</i> in der Funktion <i>Word Lookup</i> .	126
Abbildung 38: <i>Hemero</i> – Suche nach <i>naturaleza</i> und ‚sehr wenig ähnlichen Lexemen‘.	128
Abbildung 39: <i>Macrocorpus</i> – Ausschnitt aus einem Text aus Buenos Aires.	130
Abbildung 40: C-ORAL-ROM – Ausschnitt eines informellen, privaten Dialogs.	133
Abbildung 41: CORLEC – Beispiele für Etikettierungen.	134
Abbildung 42: CORLEC – Beispieltext zur Veranschaulichung der Transkription.	135
Abbildung 43: C-Or-DiAL – Transkription eines Dialogs zu den Themen Familie, Zukunft, Vorlieben und Jugend.	137
Abbildung 44: <i>Sacodeyl</i> – Anzeige der Konkordanzen für <i>gustaría</i> .	139
Abbildung 45: <i>Backbone Spanish</i> – Suche nach dem Substantiv <i>deporte</i> im Bereich <i>Co-Occurrence</i> .	140
Abbildung 46: SPT – Fragen zur Grammatik in einem Lernervideo.	141
Abbildung 47: SPT – Transkript eines Lernervideos.	142
Abbildung 48: CORPES – Flexionsmorphologische Informationen zum Lemma <i>libro</i> in der Form <i>libros</i> .	143
Abbildung 49: CT-IULA – Suche nach dem Lemma <i>contaminación</i> im Unterbereich <i>Ciencia y tecnología ambiental y energética</i> .	144

Abbildung 50: <i>El Grial</i> – Ergebnisanzeige bei der Suche nach <i>televisión</i> in der Option <i>Búsqueda Simple</i> .	145
Abbildung 51: <i>El Grial</i> – Benutzeroberfläche der Option <i>Búsqueda Compleja</i> .	146
Abbildung 52: <i>El Grial</i> – Option <i>El Manchador de Textos</i> bei der Suche nach <i>para</i> + <i>infinitivo</i> und <i>para que</i> + <i>subjuntivo</i> .	146
Abbildung 53: <i>El Grial</i> – Ausschnitt eines Texts in der Option <i>Visualización Output</i> .	147
Abbildung 54: BDS – Suche nach Verben, die mit der Präposition <i>hasta</i> verwendet werden.	148
Abbildung 55: <i>Europarl-es</i> – Konkordanzen des Substantivs <i>ojo</i> mit zusätzlicher Information zu Wortart und Funktion im Satz.	149
Abbildung 56: <i>Europarl-es</i> – Analyse eines Textausschnitts.	150
Abbildung 57: AnCora – Beispiel für die Darstellung eines Teilsatzes als Baumdiagramm.	151
Abbildung 58: <i>SenSem</i> – Suche nach dem Verb <i>contar</i> in der Bedeutung <i>ser significativo o relevante</i> .	153
Abbildung 59: <i>Internet with semantic roles</i> – Ergebnisse der Suche nach dem Substantiv <i>profesor</i> als <i>agent</i> .	154
Abbildung 60: <i>Internet with semantic roles</i> – Ergebnisse der Suche nach dem Substantiv <i>profesor</i> als <i>speaker</i> .	154
Abbildung 61: <i>Internet with semantic roles</i> – Ergebnisse der Suche nach dem Substantiv <i>profesor</i> als <i>patient</i> .	154
Abbildung 62: CdE:Nuevo – Suche nach Synonymen für das Lemma <i>guapo</i> .	160
Abbildung 63: CdE:Nuevo – Beginn der Übersicht der Synonyme für das Lemma <i>guapo</i> .	161
Abbildung 64: CdE:Nuevo – Übersicht der Synonyme für das Lemma <i>guapo</i> im Vergleich der Varietäten.	161
Abbildung 65: <i>WordAndPhrase</i> – Eintrag des Substantivs <i>problema</i> .	162
Abbildung 66: <i>Google Books: Spanish</i> – Suche nach Kollokationen mit dem Substantiv <i>nariz</i> und einem Adjektiv.	164
Abbildung 67: <i>NOW Corpus: Spanish</i> – Vergleich der Verwendung von <i>NOUN</i> + <i>dulce</i> in Spanien und Mexiko.	166
Abbildung 68: <i>NOW Corpus: Spanish</i> – Vergleich der Substantive mit der Endung <i>-ismo</i> von 2012-2015 mit denjenigen von 2016-2018.	166
Abbildung 69: <i>esTenTen</i> – Kollokationen zum Substantiv <i>pregunta</i> .	171
Abbildung 70: <i>esTenTen</i> – Ähnliche Lexeme für das Substantiv <i>pregunta</i> .	172
Abbildung 71: ADMYTE – Auszug aus dem Korpus ADMYTE.	173
Abbildung 72: PRESEEA – Ausschnitt aus einem Interview mit einer Frau der oberen Bildungsschicht aus Madrid.	176
Abbildung 73: LEXESP – Suche nach dem Lemma <i>escribir</i> .	177
Abbildung 74: <i>WebCorp Live</i> – Benutzeroberfläche in der Grundeinstellung.	187

7. Abbildungsverzeichnis

Abbildung 75: Schema zur Veranschaulichung der Auswirkungen von <i>Herausforderung</i> und <i>Unterstützung</i> auf den Lernprozess nach Gilmore (2007).	207
Abbildung 76: Ausschnitt aus dem <i>Diccionario combinatorio práctico del español contemporáneo</i> .	276
Abbildung 77: Die Rangplätze 2000 bis 2019 im <i>Routledge Frequency Dictionary of Spanish</i> .	280
Abbildung 78: Anzahl der in Lehrwerken unterrepräsentierten Lexeme je Wortart....	304
Abbildung 79: Schema zur Abgrenzung von Kollokationen nach Pöll (2002).	331
Abbildung 80: CORPES – <i>Coapariciones</i> des Substantivs <i>error</i> , Ergebnisse 1-20. ...	348
Abbildung 81: CORPES – <i>Coapariciones</i> des Substantivs <i>error</i> , Ergebnisse 21-40. .	348
Abbildung 82: CdE:Orig (Version 2008) – Suche nach <i>bella</i> mit einem beliebigen Substantiv.	349
Abbildung 83: CdE:Orig (Version 2008) – Ergebnis der Suche nach Substantiven, die zusammen mit <i>bella</i> auftreten.	350
Abbildung 84: CdE:Orig (Version 2008) – Vergleich der Adjektive im Umfeld von <i>mujer</i> und <i>hombre</i> .	351
Abbildung 85: CdE:Orig (Version 2008) – Ergebnis des Vergleichs der Adjektive im Umfeld von <i>mujer</i> und <i>hombre</i> .	351
Abbildung 86: CORPES – Suche nach Kollokationen mit dem Substantiv <i>problema</i> .	353
Abbildung 87: CORPES – Auflistung der ersten 20 <i>Coapariciones</i> des Verbs <i>tener</i> angeordnet nach dem Wert der MI.	357
Abbildung 88: CdE:Orig (Version 2008) – Einstellungen für die Suche nach Kollokationen bestehend aus dem Substantiv <i>pelo</i> und einem Adjektiv.	359
Abbildung 89: CdE:Orig (Version 2008) – Ergebnisse der Suche nach Kollokationen von <i>pelo</i> in Verbindung mit einem Adjektiv.	359
Abbildung 90: Mindmap zum Thema ‚ <i>pelo</i> ‘.	360
Abbildung 91: <i>Spa_newscrawl_2011</i> – Graph zum Substantiv <i>cielo</i> .	361
Abbildung 92: <i>NOW Corpus: Spanish</i> – Liste der mit dem Substantiv <i>viento</i> erscheinenden Substantive in einem selbst erstellten virtuellen Korpus mit dem Titel ‚ <i>viento</i> ‘.	362
Abbildung 93: <i>NOW Corpus: Spanish</i> – Liste der mit dem Substantiv <i>lluvia</i> erscheinenden Adjektive in einem selbst erstellten virtuellen Korpus mit dem Titel ‚ <i>lluvia</i> ‘.	363
Abbildung 94: CORPES – <i>Coapariciones</i> des Substantivs <i>pelo</i> .	364
Abbildung 95: Kollokationsfeld zum Substantiv ‚ <i>pelo</i> ‘.	365
Abbildung 96: Kollokationsfeld zum Substantiv ‚ <i>pelo</i> ‘ inklusive der Angabe der deutschen Entsprechung.	365
Abbildung 97: <i>Google Books: Spanish</i> – Kollokationen mit dem Substantiv <i>cielo</i> und einem Adjektiv.	368
Abbildung 98: CORPES – Suche nach Adjektiven im Umfeld von <i>problema</i> .	371
Abbildung 99: CdE:Orig – Möglichkeiten der Suche nach Synonymen.	382

Abbildung 100: CdE:Orig (Version 2008) – Suche nach Substantiven, die häufig mit <i>guapo</i> oder <i>bonito</i> kombiniert werden.	383
Abbildung 101: CdE:Orig – Suche nach Synonymen von <i>barba</i>	384
Abbildung 102: CdE:Orig – Suche nach Synonymen von <i>decir</i>	384
Abbildung 103: Spanisches Wortfeld zum deutschen Begriff ‚sagen‘.	385
Abbildung 104: CdE:Orig – Suche nach Synonymen von <i>problema</i>	385
Abbildung 105: CdE:Orig (Version 2008) – Suche nach Synonymen von <i>grande</i> , die zusammen mit dem Substantiv <i>ventaja</i> verwendet werden.	387
Abbildung 106: CdE:Nuevo – Suche nach <i>cabeza</i> mit unterschiedlichen Artikeln.	392
Abbildung 107: CdE:Nuevo – Konkordanzen des Substantivs <i>cabeza</i> mit definitivem, femininem Artikel.	392
Abbildung 108: CdE:Nuevo – Konkordanzen des Substantivs <i>cabeza</i> mit definitivem, maskulinem Artikel.	392
Abbildung 109: CdE:Nuevo – Konkordanzen für das Substantiv <i>estrella</i>	396
Abbildung 110: CdE:Orig – Möglichkeiten der Suche mit Platzhaltern.	397
Abbildung 111: CdE:Orig – Auflistung der Ergebnisse zum Suchausdruck <i>familia*</i>	400
Abbildung 112: CORPES – Suche nach der 1. Person Singular von <i>conocer</i> im <i>presente de indicativo</i>	411
Abbildung 113: CORPES – Suche nach der 1. Person Singular von <i>continuar</i> im <i>pretérito indefinido</i>	413
Abbildung 114: CORPES – Suche nach der Form des <i>participio de pasado</i> von <i>hablar</i>	414
Abbildung 115: CORPES – Suche nach der 3. Person Plural des Verbs <i>traer</i> im <i>pretérito indefinido</i>	415
Abbildung 116: Auflistung der <i>verbos de juicio y expresiones impersonales</i> des Lehrwerks <i>¡Adelante! Nivel intermedio</i>	425
Abbildung 117: CORPES – Suche nach dem Lexem <i>México</i> im mündlichen Subkorpus.	430
Abbildung 118: <i>WordAndPhrase</i> – Bearbeitung eines spanischen Texts zum Fest <i>Inti Raymi</i>	443
Abbildung 119: <i>WordAndPhrase</i> – Eintrag des Verbs <i>repartir</i>	444
Abbildung 120: Vokabelerklärung für vier unbekannte Lexeme aus dem Text <i>Inti Raymi</i>	445
Abbildung 121: Lückenübung für vier unbekannte Lexeme aus dem Text <i>Inti Raymi</i>	445
Abbildung 122: COLA – Ausschnitt aus einem Gespräch zwischen drei Jugendlichen über einen bevorstehenden Physiktest.	450
Abbildung 123: COLA – Ausschnitt aus einem Gespräch zwischen einem Jungen und einem Mädchen auf der Straße.	450
Abbildung 124: CORPES – Suche nach der Verbform der 2. Person Singular von <i>vivir</i> im <i>voseo</i>	453

7. Abbildungsverzeichnis

Abbildung 125: CORPES – Suche nach <i>mayo</i> in gesprochenen Texten aus Spanien und Lateinamerika.	454
Abbildung 126: <i>AntConc</i> – Ergebnis der Suche nach <i>trabajo</i>	490
Abbildung 127: <i>AntConc</i> – Reiter <i>Concordance Plot</i> bei der Suche nach <i>inca*</i>	492
Abbildung 128: <i>AntConc</i> – <i>File View</i> bei der Suche nach <i>hambre</i> in einem ausgewählten Text.	494
Abbildung 129: <i>AntConc</i> – Funktion <i>Clusters</i> bei der Suche nach <i>N-Grams</i>	496
Abbildung 130: <i>AntConc</i> – Funktion <i>Collocates</i> bei der Suche nach <i>inca*</i>	497
Abbildung 131: <i>AntConc</i> – <i>Word List</i> des <i>Argentina</i> -Korpus unter Verwendung einer Lemmaliste.....	499
Abbildung 132: <i>Argentina</i> -Korpus – Ermittlung einer <i>Keyword List</i> unter Verwendung einer Stopliste.....	502
Abbildung 133: Parameter zur Textauswahl bei repräsentativen Korpora nach Biber (1994).	593
Abbildung 134: Herkunft der Texte des CdE:Orig (Gesamt) nach Jahrhundert.	607
Abbildung 135: Zusammensetzung des CdE:Orig nach Genres und Registern.	609
Abbildung 136: Anzahl der Einträge des CORDE in den verschiedenen Genres und Epochen.....	610

8. Anhang

1. Fragebogen: Derzeitige Nutzung linguistischer Korpora im Spanischunterricht

Fragebogen: Derzeitige Nutzung linguistischer Korpora im Spanischunterricht

Sehr geehrte Spanischlehrerinnen und Spanischlehrer,

ich bin derzeit als Lehrerin für die Fächer Englisch und Spanisch am Walter-Gropius-Gymnasium in Selb tätig und promoviere an der Universität Regensburg auf dem Gebiet der spanischen Sprachwissenschaft. Im Rahmen meiner Promotion „Didaktische Exploration linguistischer Korpora am Beispiel des Spanischunterrichts“ wäre es sehr wünschenswert, über Daten bezüglich der derzeitigen Nutzung spanischer Korpora für den Unterricht zu verfügen.

In der Sprachwissenschaft versteht man unter einem Korpus eine Sammlung von Texten, von denen angenommen wird, dass sie für eine gegebene Sprache repräsentativ sind, so dass sie für sprachwissenschaftliche Untersuchungen verwendet werden können.¹ Selbstverständlich sind digitale Korpora besonders interessant für die Verwendung im schulischen Bereich, da diese besonders umfangreich und meist frei zugänglich sind. Beispiele hierfür sind die Korpora der Real Academia Española: CREA, CORDE und CORPES.

Um nun herauszufinden, inwieweit Korpora derzeit im Unterricht oder für die Unterrichtsvorbereitung genutzt werden, möchte ich Sie bitten, den vorliegenden anonymen Fragebogen wahrheitsgemäß auszufüllen und ihn mir zukommen zu lassen. Auch wenn Sie bisher noch keinerlei Erfahrung mit Korpora jeglicher Art gemacht haben, wäre es sehr hilfreich, wenn Sie den Fragebogen ausfüllen und an mich zurückschicken könnten, da ich mit dieser Umfrage ermitteln möchte, ob und wie weit die Arbeit mit Korpora bisher verbreitet ist.

Vielen Dank schon einmal im Voraus dafür, dass Sie sich die Zeit nehmen, meine Fragen zu beantworten. Bitte schicken Sie den ausgefüllten Fragebogen per E-Mail an mich, Nadine Meier (E-Mail: nadine.nm.meier@gmail.com), zurück.

¹ Vgl. Tognini-Bonelli, Elena (2001): *Corpus Linguistics at Work*. Amsterdam and Philadelphia: John Benjamins, S. 2.

1. Nutzung spanischer Korpora für die Vor- und Nachbereitung des eigenen Unterrichts

In diesem Abschnitt wird nach der Verwendung der Korpora durch den Lehrer bei der Vor- und Nachbereitung des Unterrichts gefragt, also beispielsweise für die Erstellung von Unterrichtsmaterial, die Suche nach Beispielsätzen oder die Korrektur von Texten. Es wird also ausschließlich nach der Nutzung der Korpora durch die Lehrkraft und nicht durch die Schüler gefragt.

1.1. Benutzen Sie Korpora für die Vor- und Nachbereitung Ihres Unterrichts?

- ☐ Ja [bitte beantworten Sie auch die Fragen 1.2 - 1.6]
☐ Nein [bitte fahren Sie mit Frage 1.7 fort]

1.2. Welche Korpora haben Sie bisher dafür verwendet?

- ☐ Corpus de Referencia del Español Actual (CREA)
☐ Corpus Diacrónico del Español (CORDE)
☐ Corpus del Español del Siglo XXI (CORPES)
☐ Corpus del Español von Mark Davies (CDE)
☐ C-ORAL-ROM
☐ Sonstige:

1.3. Wie häufig verwenden Sie Korpora zur Vor- und Nachbereitung des Unterrichts?

- ☐ jeden Tag
☐ 1x pro Woche
☐ 1x pro Monat
☐ 1x pro Jahr

1.4. Für welche Themengebiete bzw. Phasen der Vor- und Nachbereitung des Unterrichts haben Sie bisher Korpora verwendet? Bitte geben Sie auch das Lernjahr und das Alter der Schüler an.

1.5. Wie würden Sie Ihre bisherigen Erfahrungen bei der Verwendung spanischer Korpora für die Vor- und Nachbereitung des Unterrichts bewerten? Beurteilen Sie Ihre Erfahrungen auf einer Skala von 1 bis 6 entsprechend Schulnoten:

sehr gut 1 2 3 4 5 6 ungenügend
☐ ☐ ☐ ☐ ☐ ☐

1.6. Welche Vor- und Nachteile sehen Sie bei der Verwendung spanischer Korpora für die Vor- und Nachbereitung des Unterrichts?

Fahren Sie nun mit Abschnitt 2 „Nutzung linguistischer Korpora im Unterricht“ fort.

1.7. Warum haben Sie bisher bei der Vor- und Nachbereitung des Unterrichts auf Korpora verzichtet?

- ☐ Mangelnde Kenntnis der existierenden Korpora
☐ Mangelnde Kenntnis der Funktionsweise von Korpora
☐ Mangelnde Zeit für die Einarbeitung in die Korpus-thematik
☐ Mangelnde Zeit für die Arbeit mit Korpora
☐ Sonstiges:

1.8. Könnten Sie sich vorstellen, hierbei in Zukunft Korpora zu nutzen, wenn es dafür einen Leitfaden gäbe?

Sehr
wahrscheinlich 1 2 3 4 5 6 Sehr
unwahrscheinlich

2. Nutzung linguistischer Korpora im Unterricht

Dieser Abschnitt umfasst Fragen zur Verwendung der Korpora im Unterricht sowohl durch Lehrer als auch durch Schüler.

2.1. Haben Sie schon einmal im Unterricht mit Korpora gearbeitet?

- ☐ Ja [bitte beantworten Sie auch die Fragen 2.2 – 2.7]
☐ Nein [bitte fahren Sie mit Frage 2.8 fort]

2.2. Mit welchen Korpora haben Sie bisher im Unterricht gearbeitet?

- ☐ Corpus de Referencia del Español Actual (CREA)
☐ Corpus Diacrónico del Español (CORDE)
☐ Corpus del Español del Siglo XXI (CORPES)
☐ Corpus del Español von Mark Davies (CDE)
☐ C-ORAL-ROM
☐ Sonstige: _____

2.3. Wie häufig arbeiten Sie im Unterricht mit Korpora?

- ☐ jeden Tag
☐ 1x pro Woche
☐ 1x pro Monat
☐ 1x pro Jahr

2.4. Bei welchen Themengebieten bzw. Unterrichtseinheiten haben Sie bisher im Unterricht mit Korpora gearbeitet? Bitte geben Sie auch das Lernjahr und das Alter der Schüler an.

2.5. Wie würden Sie Ihre bisherigen Erfahrungen bei der Arbeit mit spanischen Korpora im Unterricht bewerten (auf einer Skala von 1 bis 6 entsprechend Schulnoten)?

sehr gut 1 2 3 4 5 6 ungenügend
☐ ☐ ☐ ☐ ☐ ☐

2.6. Wie würden Sie die Reaktion der Schüler auf die Arbeit mit Korpora im Unterricht bewerten (auf einer Skala von 1 bis 6 entsprechend Schulnoten)?

sehr gut 1 2 3 4 5 6 ungenügend
☐ ☐ ☐ ☐ ☐ ☐

2.7. Welche Vor- und Nachteile sehen Sie bei der Arbeit mit spanischen Korpora im Unterricht?

Fahren Sie nun mit Abschnitt 3 „Erstellung und Nutzung eines eigenen Korpus“ fort.

2.8. Warum haben Sie bisher im Unterricht auf Korpora verzichtet?

- ☐ Mangelnde Kenntnis der existierenden Korpora
☐ Mangelnde Kenntnis der Funktionsweise von Korpora

Ihrerseits

- ☐ Mangelnde Kenntnis der Funktionsweise von Korpora seitens der Schüler
☐ Mangelnde Zeit für die Einarbeitung in die Korpus-thematik Ihrerseits
☐ Mangelnde Zeit für die Einarbeitung in die Korpus-thematik seitens der Schüler
☐ Mangelnde Zeit für die Arbeit mit Korpora Ihrerseits
☐ Mangelnde Unterrichtszeit
☐ Sonstiges: _____

2.9. Könnten Sie sich vorstellen, in Zukunft im Unterricht mit Korpora zu arbeiten, wenn es dafür einen Leitfaden gäbe?

Sehr
wahrscheinlich 1 2 3 4 5 6 Sehr
unwahrscheinlich

3. Erstellung und Nutzung eines eigenen Korpus

In diesem Abschnitt wird nach der Erstellung und Nutzung eines eigenen Korpus durch die Lehrkraft (mit und ohne Schüler) gefragt. Es handelt sich hier nicht um die Verwendung bereits existierender Korpora.

3.1. Haben Sie schon einmal ein eigenes Korpus erstellt?

- ☐ Ja [bitte beantworten Sie auch die Fragen 3.2 – 3.12]
☐ Nein [bitte fahren Sie mit Frage 3.13 fort]

3.2. Welche Art Korpus haben Sie erstellt? Mehrere Antworten sind möglich.

- ☐ Korpus der geschriebenen Sprache
☐ Korpus der gesprochenen Sprache
☐ Korpus der Gegenwartssprache
☐ Historisches Korpus
☐ Monolinguales Korpus
☐ Multilinguales Korpus
☐ Sonstiges: _____

3.3. Wie groß ist das von Ihnen erstellte Korpus (Anzahl in Wörtern)?

3.4. Welche Programme haben Sie für die Erstellung Ihres Korpus verwendet? Ist es ein privates Korpus oder ist es öffentlich zugänglich?

3.5. Wofür haben Sie das Korpus genutzt?

Bitte geben Sie auch das Lernjahr und das Alter der Schüler an, sofern Sie das Korpus zusammen mit Ihren Schülern genutzt haben.

3.6. Haben Sie das Korpus allein erstellt, zusammen mit anderen Lehrern oder zusammen mit Ihren Schülern?

- ☐ allein
☐ zusammen mit anderen Lehrern
☐ zusammen mit Schülern
☐ Sonstiges: _____

3.7. Haben Sie das Korpus allein genutzt oder zusammen mit Ihren Schülern?

- ☐ allein
☐ zusammen mit Schülern
☐ Sonstiges:

3.8. Wie würden Sie die Erstellung dieses Korpus bewerten (auf einer Skala von 1 bis 6 entsprechend Schulnoten)?

sehr gut 1 2 3 4 5 6 ungenügend
☐ ☐ ☐ ☐ ☐ ☐

3.9. Wie würden Sie die Nutzungsmöglichkeiten dieses Korpus bewerten (auf einer Skala von 1 bis 6 entsprechend Schulnoten)?

sehr gut 1 2 3 4 5 6 ungenügend
☐ ☐ ☐ ☐ ☐ ☐

3.10. Falls zutreffend: Wie waren die Reaktionen der Schüler auf die Erstellung des Korpus? Beurteilen Sie die Reaktion auf einer Skala von 1 bis 6 entsprechend Schulnoten:

sehr gut 1 2 3 4 5 6 ungenügend
☐ ☐ ☐ ☐ ☐ ☐

3.11. Falls zutreffend: Wie waren die Reaktionen der Schüler auf die Nutzung des Korpus? Beurteilen Sie die Reaktion auf einer Skala von 1 bis 6 entsprechend Schulnoten:

sehr gut 1 2 3 4 5 6 ungenügend
☐ ☐ ☐ ☐ ☐ ☐

3.12. Welche Vor- und Nachteile sehen Sie in der Erstellung eines eigenen Korpus?

Fahren Sie nun mit Abschnitt 4 „Angaben zur eigenen Person“ fort.

3.13. Aus welchem Grund haben Sie bisher kein Korpus erstellt?

- ☐ Mangelndes Wissen über die Erstellung eines Korpus
☐ Mangelnde Zeit für die Erstellung eines Korpus
☐ Mangelnde Unterrichtszeit
☐ Sonstiges:

3.14. Könnten Sie sich vorstellen, einmal ein eigenes Korpus zu erstellen, wenn es dafür einen Leitfaden gäbe?

Sehr 1 2 3 4 5 6 Sehr
 wahrscheinlich ☐ ☐ ☐ ☐ ☐ ☐ unwahrscheinlich

4. Angaben zur eigenen Person

Bitte beantworten Sie nun noch ein paar Fragen zu Ihrer Person, um eine Auswertung der Ergebnisse nach Bundesland, Schultart, Altersgruppe,

Geschlecht und genereller Computernutzung zu ermöglichen.

Selbstverständlich bleibt die Anonymität der Umfrage trotz dieser Angaben aufrechterhalten.

4.1. In welchem Bundesland unterrichten Sie?

- ☐ Baden-Württemberg
☐ Bayern
☐ Berlin
☐ Brandenburg
☐ Bremen
☐ Hamburg
☐ Hessen
☐ Mecklenburg-Vorpommern
☐ Niedersachsen
☐ Nordrhein-Westfalen
☐ Rheinland-Pfalz
☐ Saarland
☐ Sachsen
☐ Sachsen-Anhalt
☐ Schleswig-Holstein
☐ Thüringen
☐ Sonstiges:

4.2. An welcher der genannten Schulen unterrichten Sie?

- ☐ Gymnasium
☐ Integrierte Gesamtschule
☐ Realschule
☐ Berufsschule (Berufsfachschule, Fachgymnasium)
☐ Volkshochschule
☐ Sonstiges:

4.3. Zu welcher der genannten Altersgruppen zählen Sie?

- ☐ 24-35 Jahre ☐ 36-45 Jahre
☐ 46-55 Jahre ☐ 56-65 Jahre
☐ Sonstiges:

4.4. Welchem Geschlecht gehören Sie an?

- ☐ männlich ☐ weiblich
☐ Sonstiges:

4.5. Wie hoch ist der Anteil der Unterrichts-vorbereitungszeit, in der Sie den Computer nutzen?

- ☐ 0-10% der Unterrichtsvorbereitung
☐ 11-30% der Unterrichtsvorbereitung
☐ 31-50% der Unterrichtsvorbereitung
☐ 51-70% der Unterrichtsvorbereitung
☐ 71-90% der Unterrichtsvorbereitung
☐ 91-100% der Unterrichtsvorbereitung

4.6. Wie hoch ist der Anteil der Unterrichtszeit, in der Sie mit den Schülern den Computer nutzen?

- ☐ 0-10% der Unterrichtszeit
☐ 11-30% der Unterrichtszeit
☐ 31-50% der Unterrichtszeit
☐ 51-70% der Unterrichtszeit
☐ 71-90% der Unterrichtszeit
☐ 91-100% der Unterrichtszeit

5. Kommentare und Anregungen

5. Sollten Sie noch weitere Kommentare, Anregungen oder Fragen zum Thema „Nutzung linguistischer Korpora im Spanischunterricht“ haben, können Sie mir diese hier gerne mitteilen:



Vielen herzlichen Dank dafür, dass Sie sich die Zeit genommen haben, die Fragen nach Ihrer persönlichen Nutzung linguistischer Korpora zu beantworten. Sie haben mir dadurch sehr dabei geholfen, die Basis für mein Promotionsprojekt zu schaffen.

Mit freundlichen Grüßen
Nadine Meier (StRin)

2. Ergebnisse des Fragebogens zur derzeitigen Nutzung linguistischer Korpora im Spanischunterricht

1. Nutzung spanischer Korpora für die Vor- und Nachbereitung des eigenen Unterrichts

1.1. Benutzen Sie Korpora für die Vor- und Nachbereitung Ihres Unterrichts?

Ja	Nein
0	32

1.2. Welche Korpora haben Sie bisher dafür verwendet?

CREA	CORDE	CORPES	CDE	C-ORAL	Sonstige:
-	-	-	-	-	-

1.3. Wie häufig verwenden Sie Korpora für die Vor- und Nachbereitung Ihres Unterrichts?

jeden Tag	1x pro Woche	1x pro Monat	1x pro Jahr
-	-	-	-

1.4. Für welche Themengebiete bzw. Phasen der Vor- und Nachbereitung des Unterrichts haben Sie bisher Korpora verwendet?

-

1.5. Wie würden Sie Ihre bisherigen Erfahrungen bei der Verwendung spanischer Korpora für die Vor- und Nachbereitung des Unterrichts bewerten?

-

1.6. Welche Vor- und Nachteile sehen Sie bei Verwendung spanischer Korpora für die Vor- und Nachbereitung des Unterrichts?

-

1.7. Warum haben Sie bisher bei der Vor- und Nachbereitung des Unterrichts auf Korpora verzichtet?

Grund	Anzahl	Prozentsatz
Mangelnde Kenntnis der existierenden Korpora	29	90,63
Mangelnde Kenntnis der Funktionsweise von Korpora	14	43,75
Mangelnde Zeit für die Einarbeitung in die Korpusthematik	6	18,75
Mangelnde Zeit für die Arbeit mit Korpora	3	9,38
Sonstiges	5	15,63

Sonstiges:

(A) Wenn ich literarische Texte verwende, weiß ich nicht, ob sie unter den o.g. Quellen auftauchen. Wenn das so ist, schaue ich aber nicht extra dort nach. Ich benutze Internet-Artikel für meine Unterrichtsvorbereitung und vereinfache sie. Ich habe anhand Ihrer Erklärung nicht verstanden, ob solche Texte unter das fallen, was Sie unter "Korpus" verstehen?

(B) Ich bin derzeit im 2. Ausbildungsabschnitt und unterrichte kein Spanisch (Erweiterungsfach). Im 1. AA hatte ich eine 8. Klasse, da wäre die Nutzung spanischer Korpora gar nicht relevant gewesen. Doch selbst wenn es eine Oberstufe gewesen wäre – ich kenne mich überhaupt nicht damit aus und hatte Korpora auch gar nicht “auf dem Schirm”. Hängt wohl auch mit meinem Erweiterungs-Studium zusammen, das sich im Wesentlichen auf das Examen beschränkt hat. Als Hauptgrund generell wäre also wohl einfach Ahnungslosigkeit zu nennen.

(C) Ich kenne das nicht.

(D) keine Notwendigkeit

(E) -

- 1.8. Könnten Sie sich vorstellen, hierbei in Zukunft Korpora zu nutzen, wenn es dafür einen Leitfaden gäbe?

sehr wahrscheinlich	1	2	3	4	5	6	sehr unwahrscheinlich
	6	9	9	3	2	3	Ø 2,84

2. Nutzung linguistischer Korpora im Unterricht

- 2.1. Haben Sie schon einmal im Unterricht mit Korpora gearbeitet?

Ja	Nein
0	32

- 2.2. Mit welchen Korpora haben Sie bisher im Unterricht gearbeitet?

CREA	CORDE	CORPES	CDE	C-ORAL	Sonstige:
-	-	-	-	-	-

- 2.3. Wie häufig arbeiten Sie im Unterricht mit Korpora?

jeden Tag	1x pro Woche	1x pro Monat	1x pro Jahr
-	-	-	-

- 2.4. Bei welchen Themengebieten bzw. Unterrichtseinheiten haben Sie bisher im Unterricht mit Korpora gearbeitet?

-

- 2.5. Wie würden Sie Ihre bisherigen Erfahrungen bei der Arbeit mit spanischen Korpora im Unterricht bewerten?

-

- 2.6. Wie würden Sie die Reaktion der Schüler auf die Arbeit mit Korpora im Unterricht bewerten?

-

- 2.7. Welche Vor- und Nachteile sehen Sie bei der Arbeit mit spanischen Korpora im Unterricht?

-

- 2.8. Warum haben Sie bisher im Unterricht auf Korpora verzichtet?

Grund	Anzahl	Prozentsatz
Mangelnde Kenntnis der existierenden Korpora	30	93,75
Mangelnde Kenntnis der Funktionsweise von Korpora Ihrerseits	18	56,25
Mangelnde Kenntnis der Funktionsweise von Korpora seitens der Schüler	8	25
Mangelnde Zeit für die Einarbeitung in die Korpus thematik Ihrerseits	6	18,75
Mangelnde Zeit für die Einarbeitung in die Korpus thematik seitens der Schüler	2	6,25
Mangelnde Zeit für die Arbeit mit Korpora Ihrerseits	4	12,5
Mangelnde Unterrichtszeit	4	12,5
Sonstiges	1	3,13

Sonstiges:

(A) keine Notwendigkeit

- 2.9. Könnten Sie sich vorstellen, hierbei in Zukunft Korpora zu nutzen, wenn es dafür einen Leitfaden gäbe?

sehr wahrscheinlich	1	2	3	4	5	6	sehr unwahrscheinlich
	3	8	6	7	3	3	Ø 3,27

3. Erstellung und Nutzung eines eigenen Korpus

- 3.1. Haben Sie schon einmal ein eigenes Korpus erstellt?

Ja	Nein
0	32

- 3.2. Welche Art Korpus haben Sie erstellt?

-

- 3.3. Wie groß ist das von Ihnen erstellte Korpus?

-

- 3.4. Welche Programme haben Sie für die Erstellung Ihres Korpus verwendet? Ist es ein privates Korpus oder ist es öffentlich zugänglich?

-

- 3.5. Wofür haben Sie das Korpus genutzt?

-

8. Anhang

3.6. Haben Sie das Korpus allein erstellt, zusammen mit anderen Lehrern oder zusammen mit Ihren Schülern?

-

3.7. Haben Sie das Korpus allein genutzt oder zusammen mit Ihren Schülern?

-

3.8. Wie würden Sie die Erstellung dieses Korpus bewerten?

-

3.9. Wie würden Sie die Nutzungsmöglichkeiten dieses Korpus bewerten?

-

3.10. Falls zutreffend: Wie waren die Reaktionen der Schüler auf die Erstellung des Korpus?

-

3.11. Falls zutreffend: Wie waren die Reaktionen der Schüler auf die Nutzung des Korpus?

-

3.12. Welche Vor- und Nachteile sehen Sie in der Erstellung eines eigenen Korpus?

-

3.13. Aus welchem Grund haben Sie bisher kein Korpus erstellt?

Grund	Anzahl	Prozentsatz
Mangelndes Wissen über die Erstellung eines Korpus	31	96,88
Mangelnde Zeit für die Erstellung eines Korpus	13	40,63
Mangelnde Unterrichtszeit	6	18,75
Sonstiges	2	6,25

Sonstiges:

(A) Ich weiß gar nicht so recht, was damit gemeint ist – wie sollte man so etwas erstellen? Klingt nach Doktorarbeit-Umfang.

(B) -

3.14. Könnten Sie sich vorstellen, einmal ein eigenes Korpus zu erstellen, wenn es dafür einen Leitfaden gäbe?

sehr wahrscheinlich	1	2	3	4	5	6	sehr unwahrscheinlich
	1	7	9	4	5	5	Ø 3,65

4. Angaben zur eigenen Person

4.1. In welchem Bundesland unterrichten Sie?

Bundesland	Anzahl	Prozentsatz
Baden-Württemberg	-	-
Bayern	15	46,88
Berlin	1	3,13
Brandenburg	-	-
Bremen	3	9,38
Hamburg	-	-
Hessen	-	-
Mecklenburg-Vorpommern	-	-
Niedersachsen	3	9,38
Nordrhein-Westfalen	4	12,5
Rheinland-Pfalz	-	-
Saarland	-	-
Sachsen	-	-
Sachsen-Anhalt	-	-
Schleswig-Holstein	6	18,75
Thüringen	-	-
Sonstiges	-	-
	32	100

4.2. An welcher der genannten Schulen unterrichten Sie?

Schulart	Anzahl	Prozentsatz
Gymnasium	28	87,5
Integrierte Gesamtschule	-	-
Realschule	-	-
Berufsschule	2	6,25
Volkshochschule	-	-
Sonstiges	2	6,25
	32	100

Sonstiges:

(A) FOS

(B) Gemeinschaftsschule mit Oberstufe

4.3. Zu welcher der genannten Altersgruppen zählen Sie?

Altersgruppe	Anzahl	Prozentsatz
24-35 Jahre	11	34,38
36-45 Jahre	11	34,38
46-55 Jahre	6	18,75
56-65 Jahre	4	18,75
Sonstiges	-	-
	32	100

4.4. Welchem Geschlecht gehören Sie an?

Geschlecht	Anzahl	Prozentsatz
männlich	7	22,58
weiblich	24	77,42
Sonstiges	-	-
	31	100

4.5. Wie hoch ist der Anteil der Unterrichtsvorbereitung, in der Sie den Computer nutzen?

Vorbereitung mit Computer	Anzahl	Prozentsatz
0-10%	1	3,23
11-30%	1	3,23
31-50%	5	16,13
51-70%	11	35,48
71-90%	11	35,48
91-100%	2	6,45
	31	100

4.6. Wie hoch ist der Anteil der Unterrichtszeit, in der Sie mit den Schülern den Computer nutzen?

Unterrichtszeit mit Computer	Anzahl	Prozentsatz
0-10%	20	64,52
11-30%	9	29,03
31-50%	2	6,45
51-70%	-	-
71-90%	-	-
91-100%	-	-
	31	100

5. Kommentare und Anregungen

- (A) Die Verwendung digitaler Korpora ist sicher hilfreich und sinnvoll, da man sich vermutlich dadurch langes Suchen in Lexika oder Internet erspart. Bekannt war mir nach Spanischstudium und Referendariat leider keines.
- (B) Ich habe in der Einleitung ehrlich gesagt nicht verstanden, was Korpora sind und wofür ich sie verwenden kann. Die genannten Beispiele sind für mich wenig anschaulich. Vielleicht ist es unter einem gängigeren Namen bekannt? Ich hatte jedenfalls noch nie davon gehört.

- (C) Ich bin völlig ahnungslos und an neuem Input interessiert (Angabe der E-Mail-Adresse).
- (D) Ich bin im Referendariat in Bremen und habe bis zum Schluss nicht vollends verstanden, worum es in diesem Fragebogen geht. Anscheinend sind Sie sehr in der Materie drin. Sie geben eine allgemeine Definition (ohne Beispiel) und fragen dann in den Fragen nach der konkreten Verwendung (CRAE usw.). Das kann für Lehrkräfte, die lange aus der Uni raus und nicht in der Materie sind und die zumeist lehrwerkbasierten Unterricht machen, sehr abstrakt wirken. So auch für mich. Im Unterricht verwende ich die Vokabeln, die im Buch stehen. Habe bisher weder im Ref noch in meinen Praktika eine andere Vorgehensweise beobachten können. Viel Erfolg bei Ihrem Forschungsvorhaben!

3. Parameter zur Textauswahl bei repräsentativen Korpora nach Biber (1994)

Situational parameters listed as hierarchical sampling strata:	
1. Primary channel:	written / spoken / scripted speech
2. Format:	published / not published (+ various formats within published)
3. Setting:	institutional / other public / private-personal
4. Addressee:	<ul style="list-style-type: none"> a. plurality: unenumerated / plural / individual / self b. presence (place and time): present / absent c. interactiveness: none / little / extensive d. shared knowledge: general / specialized / personal
5. Addressor:	<ul style="list-style-type: none"> a. demographic variation: sex, age, occupation, etc. b. acknowledgement: acknowledged individual/ institution
6. Factuality:	factual-informational / intermediate or indeterminate / imaginative
7. Purposes:	persuade, entertain, edify, inform, instruct, explain, narrate, describe, keep records, reveal self, express attitudes, opinions, or emotions, enhance interpersonal relationship, ...
8. Topics:	...

Abbildung 133: Parameter zur Textauswahl bei repräsentativen Korpora nach Biber (1994).¹³²⁷

¹³²⁷ Vgl. Biber (1994): op. cit., S. 382.

4. Alphabetische Übersicht über die Korpora der spanischen Sprache

Name	Art des Korpus	Größe	Verfügbarkeit	Charakteristik und Einsatzmöglichkeiten
ADMYTE – <i>Archivo Digital de Manuscritos y Textos Españoles</i> (University of Wisconsin-Madison)	Historisches Korpus	290 Werke des Spanischen im Mittelalter (mehr als 54.000 Seiten)	Als CD-ROM erhältlich und im Internet gegen eine Gebühr von ca. 100 € pro Jahr nutzbar (www.admyte.com)	Viele mittelalterliche Meisterwerke;
AnCora (<i>Grupo de Investigación CLiC</i>)	Morphosyntaktisch, syntaktisch und semantisch annotiertes sowie lemmatisiertes Spezialkorpus	500.000 Lexeme (Spanisch) 500.000 Lexeme (Katalanisch)	Im Internet kostenlos verfügbar (http://clic.ub.edu/corpus/es/busquedas)	Darstellung syntaktischer Annotation im Baumdiagramm;
<i>Backbone Spanish</i> (Kurt Kohn, Universität Tübingen)	Multimodales, pädagogisches Spezialkorpus in sieben europäischen Sprachen (Deutsch, Englisch, Französisch, Polnisch Spanisch, Türkisch und Englisch als <i>lingua franca</i>)	25 spanische Interviews zwischen 10 und 20 Minuten Länge	Im Internet kostenlos verfügbar (http://webapps.ael.uni-tuebingen.de/backbone-search/faces/search.jsp)	Videsequenzen verfügbar; pädagogische Annotation; Einteilung und Suche nach Themen, Grammatik, Lexik, kommunikativen Funktionen und GER-Niveau;

8. Anhang

BDS – <i>Base de Datos Sintácticos del Español Actual</i> (Universidad de Santiago de Compostela)	Syntaktisch annotiertes Spezialkorpus	160.000 Sätze	Im Internet kostenlos verfügbar (http://www.bds.usc.es)	syntaktische Annotation (manuell);
<i>Biblia Medieval</i> (Universitat de les Illes Balears)	Historisches Korpus	5 Mio. romanische Lexeme, 1,5 Mio. lateinische und hebräische Lexeme	Im Internet kostenlos verfügbar (http://corpus.bibliamedieval.es)	Verschiedene mittelalterliche Übersetzungen der Bibel ins Kastilische;
CAES - <i>Corpus de aprendices de español como lengua extranjera</i> (Universidad de Santiago de Compostela, Instituto Cervantes)	Schriftliches Lernerkorpus (Muttersprachen: Arabisch, Chinesisch, Französisch, Englisch, Portugiesisch und Russisch)	575.000 Lexeme aus 3.878 Texten von 1.423 Studenten	Im Internet kostenlos verfügbar (http://galvan.usc.es/caes/search)	Texte von Spanischstudenten mit sechs unterschiedlichen Muttersprachen und Sprachniveaus von A1 bis C1; morphosyntaktische Annotation und Lemmatisierung;
CdE:Nuevo – <i>Corpus del Español (Web/Dialectos)</i> (Mark Davies)	Opportunistisches Korpus aus Internettexen	2 Milliarden Lexeme	Im Internet kostenlos verfügbar nach Anmeldung, allerdings auf 50 Suchanfragen pro Tag begrenzt (https://www.corpusdelespanol.org/web-dial/)	Sehr aktuell, mit Texten aus den Jahren 2013 und 2014; Zugang zu den Texten sowie offline Nutzung gegen eine Gebühr von 245-795 \$ (je nach Nutzungsart) möglich;

				Lemmatisierung, morphosyntaktische Annotation; Suche nach Synonymen möglich; Erstellung virtueller Korpora; Sehr nützlich: Programm <i>WordAndPhrase</i> ;
CdE:Orig – <i>Corpus del Español (Género/Histórico)</i> (Mark Davies)	Historisches Korpus	100 Mio. Lexeme aus über 20.000 Texten (13. bis 20. Jahrhundert)	Im Internet kostenlos verfügbar nach Anmeldung, allerdings auf 50 Suchanfragen pro Tag begrenzt (https://www.corpusdelespanol.org/hist-gen/); Alte Version von 2008: (https://www.corpusdelespanol.org/hist-gen/2008/x.asp)	Teil des aktuellen Spanisch (20 Millionen Lexeme) kann als Referenzkorpus genutzt werden; Lemmatisierung, morphosyntaktische Annotation; Suche nach Synonymen möglich; Erstellung virtueller Korpora;
CDH – <i>Corpus del Nuevo Diccionario Histórico del Español (Real Academia Española)</i>	Historisches Korpus	355 Mio. Lexeme	Im Internet kostenlos verfügbar (http://web.frl.es/CNDHE/view/inicioExterno.view)	Äußerst umfassendes Korpus; Möglichkeit, in beinahe allen Texten der Korpora der RAE gleichzeitig zu suchen;
CEA – <i>Corpus del Español Actual</i> (Carlos Subirats, Marc Ortega)	Opportunistisches Korpus	540 Mio. Lexeme	Im Internet kostenlos verfügbar (http://spanishfn.org/tools/cea/spanish)	Lemmatisierung, morphosyntaktische Annotation; Möglichkeit, sich alle Textausschnitte

8. Anhang

				mit <i>tags</i> anzeigen zu lassen;
CEDEL2 – <i>Corpus Escrito del Español L2</i> (Cristóbal Lozano)	Lernerkorpus	730.000 Lexeme von mehr als 2.300 Teilnehmern (1.700 englischsprachige Spanischlerner und 660 spanische Muttersprachler); Endversion mit 1 Mio. Lexeme geplant;	Im Internet kostenlos verfügbar (http://cedel2.learnercorpora.com)	Verfügbarkeit von Aufsätzen zu alltäglichen Themen wie der Heimatregion, einer berühmten Persönlichkeit, einem vor kurzer Zeit gesehenen Film oder dem vergangenen Urlaub;
CLUVI – <i>Corpus Lingüístico da Universidade de Vigo</i> (Xavier Gómez Guinovart)	Parallelkorpus (Spanisch, Galizisch, Baskisch, Katalanisch, Englisch, Französisch, Portugiesisch, Deutsch)	Insgesamt 25 Mio. Lexeme; Subkorpus CONSUMER umfasst über 5,5 Mio. Lexeme;	Im Internet kostenlos verfügbar (http://sli.uvigo.es/CLUVI/)	Sammlung paralleler Textkorpora;
COLA – <i>Corpus Oral de Lenguaje Adolescente</i> (Annette Myre Jørgensen)	Korpus der Jugendsprache	500.000 Lexeme	Im Internet kostenlos verfügbar (http://clu.uni.no/humfak/cola/)	Unübersichtliche, nur auf norwegisch verfügbare Benutzeroberfläche;
C-ORAL-ROM (Emanuela Cresti, Massimo Moneglia)	Vergleichskorpus der gesprochenen romanischen Sprachen	Je 300.000 Lexeme pro Sprache (Französisch,	Nur auf CD-ROM erhältlich: Cresti, Emanuela/ Moneglia, Massimo (Hrsg.) (2005): <i>C-ORAL-ROM. Integrated Reference Corpora for</i>	Tonaufnahmen und Transkriptionen aus verschiedenen Bereichen des alltäglichen Lebens

		Italienisch, Portugiesisch, Spanisch)	<i>Spoken Romance Languages</i> . Amsterdam [u.a.]: Benjamins. Die ersten 30 Sekunden der Tonaufnahmen können kostenlos im Internet angehört werden (http://www.elda.org/en/proj/coral/demo/index.html)	(formell und informell); prosodische Annotation;
CORDE – <i>Corpus Diacrónico del Español (Real Academia Española)</i>	Historisches Korpus	250 Mio. Lexeme (13. bis 20. Jahrhundert)	Im Internet kostenlos verfügbar (http://corpus.rae.es/cordenet.html)	Veraltete Texte des CREA (Monitorkorpus: enthält nur Texte der letzten 25 Jahre) gehen ins CORDE über; morphosyntaktische Annotation;
C-Or-DiAL – <i>Corpus Oral Didáctico Anotado Lingüísticamente (Laboratorio de Lingüística Italiana (Lablita) der Università degli Studi Firenze)</i>	Korpus gesprochener Sprache, didaktisches Korpus	Knapp 119.000 Lexeme, knapp 10 Stunden Sprachaufnahmen in 240 Sequenzen	Im Internet kostenlos verfügbar (http://lablita.it/app/cordial/corpus.php) Zudem ist es als CD-ROM mit gleichnamigem Buch verfügbar: Nicolás Martínez, Carlota (2012): <i>Corpus C-Or-DiAL (Corpus Oral Didáctico Anotado Lingüísticamente)</i> . Madrid: Liceus.	Didaktisches Korpus; prosodische Annotation; Angabe von neuem Vokabular und kommunikativen Funktionen;
CORELE – <i>Corpus Oral de Español como Lengua Extranjera (ELE)</i>	Lernerkorpus, Fehlerkorpus	Spanischlerner: über 55.000 Lexeme (40 Interviews, insgesamt 13	Im Internet kostenlos verfügbar (http://cartago.lllf.uam.es/corele/index.html)	Interviews von Spanischlernern mit über 9 verschiedenen Muttersprachen, inklusive Deutsch (Niveau A2 und

8. Anhang

<i>(Universidad Autónoma de Madrid)</i>		Stunden und 36 Minuten); Muttersprachler: über 9.000 Lexeme (4 Interviews, insgesamt 1 Stunde und 22 Minuten);		B1 des <i>Common European Framework of Reference</i>); Fehlerannotation, morphosyntaktische Annotation;
CORLEC – <i>Corpus de Referencia de la Lengua Española Contemporánea (Universidad Autónoma de Madrid)</i>	Korpus gesprochener Sprache	1,1 Mio. Lexeme	Im Internet kostenlos verfügbar (http://www.llf.uam.es/ESP/Corlec.html)	Keine Suchmaske vorhanden, keine Audiodateien verfügbar;
CORPES – <i>Corpus del Español del Siglo XXI (Real Academia Española)</i>	Referenzkorpus (halboffen)	Derzeit 286 Mio. Lexeme (das endgültige Korpus sollte 2018 mit 400 Mio. Lexemen veröffentlicht werden)	Im Internet kostenlos verfügbar (http://web.frl.es/CORPES/view/inicioExternoview)	Sehr aktuell, mit Texten von 2001 bis 2012; morphosyntaktische Annotation, Lemmatisierung; Tonsequenzen können angehört werden;
<i>Corpus Tècnic del IULA (Universitat Pompeu Fabra)</i>	Morphosyntaktisch annotiertes und lemmatisiertes Spezialkorpus, teilweise Parallelkorpus	33 Mio. Lexeme (Spanisch), 26 Mio. Lexeme (Katalanisch),	Im Internet kostenlos verfügbar (http://bwananet.iula.upf.edu)	Morphosyntaktische Annotation, Lemmatisierung; parallele Darstellung zweier Sprachen möglich;

		3,5 Mio. Lexeme (Englisch)		
<i>Corpus Trilingüe Paralelo GRIAL</i> (Grup de Recerca Interuniversitari en Aplicacions Lingüístiques)	Parallelkorporus (dreisprachig)	Knapp 2,3 Mio. Lexeme (Englisch, Spanisch, Katalanisch)	Im Internet kostenlos verfügbar (http://grial.uab.es/sensem/cpg)	morphosyntaktische Annotation;
CREA – <i>Corpus de Referencia del Español Actual</i> (Real Academia Española)	Referenzkorporus, Monitorkorporus	160 Mio. Lexeme	Im Internet kostenlos verfügbar: Alte Version von 2015: (http://corpus.rae.es/creanet.html) Annotierte Version: (http://web.frl.es/CREA/view/inicioExterno.view)	Einteilung in verschiedene Module: Suche nach geographischen, thematischen oder zeitlichen Charakteristika möglich; morphosyntaktische Annotation, Lemmatisierung;
<i>El Grial</i> (Pontificia Universidad Católica de Valparaíso)	Morphosyntaktisch annotiertes und lemmatisiertes Spezialkorporus des geschriebenen Spanisch	5,4 Mio. Lexeme	Im Internet kostenlos verfügbar (http://www.elgrial.cl)	morphosyntaktische Annotation, Lemmatisierung; verschiedenste Such- und Darstellungsoptionen;
<i>esTenTen</i> (SketchEngine)	Opportunistisches Webkorporus	9 Milliarden Lexeme	Im Internet gegen Gebühr verfügbar (https://www.sketchengine.eu)	morphosyntaktische Annotation, Lemmatisierung; umfassende Möglichkeit der Suche nach

8. Anhang

				Kollokationen sowie Nutzung als Thesaurus;
<i>Europarl-es</i> (VISL, <i>Corpuseye</i>)	Morphosyntaktisch und syntaktisch annotiertes Spezialkorpus	27,2 Mio. Lexeme	Im Internet kostenlos verfügbar (http://corp.hum.sdu.dk/cqp.es.html)	Morphosyntaktische und syntaktische Annotation;
<i>Fono.ele</i> (Ana Blanco Canales, María Ángeles Álvarez Martínez – <i>Universidad de Alcalá</i>)	Mündliches Lernerkorpus (Nationalitäten: Deutsch, Griechisch, Taiwanisch, Polnisch, Portugiesisch, Ägyptisch)	630 Audiodateien	Im Internet kostenlos verfügbar (http://www3.uah.es/fonoele/corpus-muestra.php)	Korpus zur Untersuchung der Phonetik des Spanischen als Fremd- oder Zweitsprache;
<i>Google Books: Spanish</i> (Mark Davies)	Opportunistisches Korpus	45 Milliarden Lexeme aus mehreren Millionen Büchern vom 19. bis 21. Jahrhundert	Im Internet kostenlos verfügbar nach Anmeldung, allerdings auf 50 Suchanfragen pro Tag begrenzt (https://googlebooks.byu.edu/x.asp)	Direkter Link zu Daten aus Google Books;
<i>Hemero – Corpus periodístico Hemero</i> (Molino de Ideas)	Spezialkorpus aus Zeitungstexten	Über 660 Mio. Lexeme	Im Internet kostenlos verfügbar (www.hemero.es)	Möglichkeit der Suche nach ‚ähnlichen‘ Lexemen; Verfügbarkeit der vollständigen Texte und Link zum ursprünglichen Zeitungsartikel;
<i>Internet with Semantic Roles</i> (VISL, <i>Corpuseye</i>)	Semantisch annotiertes Spezialkorpus	11,2 Mio. Lexeme	Im Internet kostenlos verfügbar (http://corp.hum.sdu.dk/cqp.es.html)	semantische Annotation, Suche nach semantischen Rollen möglich;

LANGSNAP – <i>Languages and Social Networks Abroad</i> (University of Southampton)	Mündliches und schriftliches Lernerkorpus (inklusive spanischer und französischer Muttersprachler)	556 Dateien von 27 Spanischlernern und 10 Muttersprachlern	Im Internet kostenlos verfügbar (http://langsnap.soton.ac.uk)	Daten zur Entwicklung der Sprachkenntnisse englischer Muttersprachler vor, während und nach einem Auslandsaufenthalt;
LEXESP – <i>Léxico informatizado del español</i> (Núria Sebastián Gallés)	Korpus geschriebener Sprache	5,5 Mio. Lexeme des gegenwärtigen Spanisch	Nur auf CD-ROM erhältlich: Sebastián Gallés, Núria/ Cuetos Vega, Fernando/ Carreiras Valiña, Manuel Francisco/ Martí Antonin, Maria Antònia (2000): <i>LEXESP. Léxico informatizado del español</i> . Barcelona: Edicions de la Universitat de Barcelona. CD-ROM. Zusätzlich gibt es zu Demonstrationszwecken eine spärlich ausgestattete Internetseite mittels derer im Korpus gesucht werden kann: (http://www.cs.upc.edu/~nlp/tools/corpus-es.php)	morphosyntaktische Annotation, Lemmatisierung;
<i>Macrocorpus de la norma lingüística culta de las principales ciudades del mundo hispánico</i> (Antonio José Samper Padilla)	Korpus des gesprochenen Spanisch	84 Stunden gesprochener spanischer Sprache (168 Sprecher aus zwölf spanischsprachigen Städten)	Nur auf CD-ROM erhältlich: Samper Padilla, Antonio José (1998): <i>Macrocorpus de la norma lingüística culta de las principales ciudades del mundo hispánico</i> . Las Palmas de Gran Canaria: Univ. de las Palmas. CD-ROM.	Interviews von einer halben Stunde Dauer mit jeweils 14 Bewohnern pro Stadt;

8. Anhang

<i>NOW Corpus: Spanish</i> (Mark Davies)	Opportunistisches, dynamisches Korpus aus online verfügbaren Zeitungen und Zeitschriften	Derzeit 7,1 Milliarden Lexeme (Juli 2019), soll pro Monat um ca. 150 Mio. Lexeme erweitert werden	Im Internet kostenlos verfügbar nach Anmeldung, allerdings auf 50 Suchanfragen pro Tag begrenzt (https://www.corpusdelespanol.org/now/)	Morphosyntaktische Annotation, Lemmatisierung; Möglichkeit zeitlicher und geographischer Vergleiche sowie Erstellung virtueller Korpora;
PRESEEA – <i>Proyecto para el Estudio Sociolingüístico del Español de España y de América</i> (Francisco Moreno Fernández)	Korpus des gesprochenen Spanisch	Korpus ist noch im Aufbau, bisher sind keine Daten zur Größe verfügbar; Interviews aus 42 spanischsprachigen Städten	Vollständiges Korpus nur auf CD-ROM erhältlich: Moreno Fernández, Francisco (2002): <i>La lengua hablada en Alcalá de Henares. Corpus PRESEEA-Alcalá</i> . Alcalá: Univ. de Alcalá. CD-ROM. Einige Interviews sind im Internet kostenlos verfügbar (http://preseea.linguas.net/Inicio.aspx)	Spontansprachliche Aufnahmen und Transkriptionen;
<i>Sacodeyl Search Tool</i> (Pascual Pérez-Paredes)	Multimodales, pädagogisches Spezialkorpus in sieben europäischen Sprachen (Deutsch, Englisch, Französisch, Italienisch, Litauisch, Rumänisch und Spanisch)	20 – 25 Videos von 10 Minuten Länge pro Sprache (25 spanische Videosequenzen)	Im Internet kostenlos verfügbar (http://sacodeyl.inf.um.es/sacodeyl-search2/)	Videosequenzen verfügbar (bislang aber nicht abrufbar); pädagogische Annotation; Einteilung und Suche nach Themen, Grammatik Lexik, kommunikativen Funktionen und GER-Niveau;
<i>SenSem (Grup de Recerca Interuniversitari en</i>	Syntaktisch-semantic annotiertes Spezialkorpus	1 Mio. Lexeme (30.000 Sätze, ausgewählt nach	Im Internet kostenlos verfügbar (http://grial.uab.es/sensem/corpus/main)	syntaktisch-semantic Annotation, Sätze werden nach Bedeutung der

<i>Aplicacions Lingüístiques)</i>		den 250 häufigsten Verben des Spanischen)		Verben geordnet und angezeigt;
<i>Spa_news_2011_3M (CLARIN FCS Aggregator)</i>	Opportunistisches Korpus mit 40 Subkorpora	Spanisches Zeitungskorpus: 3 Mio. Sätze	Im Internet kostenlos verfügbar (https://www.clarin.eu/contentsearch)	Spanisches Zeitungskorpus mit Material aus dem Jahr 2011, Lemmatisierung, aber keine Annotation;
<i>Spa_newscrawl 2011</i> (Universität Leipzig – Institut für Informatik – Abteilung ASV)	Opportunistisches Korpus	Knapp 17 Mio. Sätze, knapp 2 Mio. <i>types</i> , über 391 Mio. <i>tokens</i> ;	Im Internet kostenlos verfügbar (http://corpora.uni-leipzig.de/de?corpusId=spa_newscrawl_2011).	Spanisches Zeitungskorpus mit Material aus dem Jahr 2011 inklusive automatischer Angabe von Kookkurrenzen und eines Graphen zur Veranschaulichung derselben; weder lemmatisiert noch annotiert;
<i>Spanish in Texas Corpus</i> (Almeida Jacqueline Toribio, Barbara Bullock)	Multimodales Korpus	Über 500.000 Lexeme von 97 bilingualen Texanern	Im Internet kostenlos verfügbar nach Anmeldung (http://corpus.spanishintexas.org) Ausgewählte Videos sind im <i>SpinTX Video Archive</i> ohne vorherige Anmeldung verfügbar (http://www.coerll.utexas.edu/spintx/)	Spanische und bilingual spanisch-englische Interviews und Gespräche;
SPLLOC – <i>Spanish Learner</i>	Mündliches Lernerkorpus	Über 500 Audiodateien	Im Internet kostenlos verfügbar (http://www.splloc.soton.ac.uk)	Daten zur Aneignung der spanischen Sprache durch

8. Anhang

<i>Language Oral Corpus</i> (Laura Domínguez, Rosamond Mitchell, Florence Myles, Emma Marsden)	(inklusive Muttersprachler)	sowie deren Transkripte		englische Muttersprachler;
SPT – <i>Spanish Corpus Proficiency Level Training</i> (Dale Koike)	Multimodales Lernerkorpus (keine Muttersprachler)	Videos von 16 englischsprachigen Spanischlernern zu neun verschiedenen Themen	Im Internet kostenlos verfügbar (http://www.laits.utexas.edu/spt/intro)	Korpus für das Training von Spanischlehrern im Hinblick auf die Bewertung von Schülern;

5. CdE:Orig (Gesamt) – Herkunft der Texte nach Jahrhundert

Jahrhundert	Lexeme	Texte	Quellen
13. Jh.	6.905.000	71	<ul style="list-style-type: none"> • [HSMS] Elektronische Texte und Konkordanzen des <i>Madison Corpus of Early Spanish Manuscripts and Printings</i>. John O'Neill (Madison und New York, 1999).
14. Jh.	2.820.000	50	<ul style="list-style-type: none"> • ADMYTE (<i>Archivo Digital de Manuscritos y Textos Españoles</i>). Vol. 0 und 2.
15. Jh.	8.515.000	160	<ul style="list-style-type: none"> • <i>Biblioteca Virtual</i> • Gonzalo de Berceo: <i>Obras Completas</i>
16. Jh.	18.001.000	323	<ul style="list-style-type: none"> • <i>Biblioteca Virtual</i> [16. bis 18. Jh.]
17. Jh.	12.746.000	499	<ul style="list-style-type: none"> • COMEDIA (<i>University of Arizona</i>) [17. Jh.]
18. Jh.	10.263.000	159	<ul style="list-style-type: none"> • <i>Proyecto Filosofía en español</i> [18. Jh.]
19. Jh.	20.465.000	392 Romane	<ul style="list-style-type: none"> • <i>Biblioteca Virtual</i>
20. Jh. Wissen- schaftliche Texte	5.138.077	2931 Artikel	<ul style="list-style-type: none"> • <i>Encarta</i>
20. Jh. Presse	5.144.631	6810 Artikel	<ul style="list-style-type: none"> • <i>ABC Cultural</i> 1991-1995 • Nachrichten – Argentinien – <i>La Prensa</i> • Nachrichten – Argentinien – <i>El Cronista</i> • Nachrichten – Bolivien – <i>ERBOL</i> • Nachrichten – Peru – <i>Caretas</i> • Nachrichten – Kolumbien – <i>Semana</i> • Nachrichten – Kuba/USA – <i>CubaNet</i>

8. Anhang

20. Jh. Fiktion	5.144.073	850 Romane / Geschich- ten	<ul style="list-style-type: none"> • <i>Biblioteca Virtual</i> • [BYU] <i>Humanities Research Center, Brigham Young University</i> • [ARG] <i>Corpus lingüístico de referencia de la lengua española en Argentina</i> • [Chile] <i>Corpus lingüístico de referencia de la lengua en Chile</i> • <i>Proyecto Sherezade</i> • Badosa.com • Ficticiosa.com
20. Jh. Mündliche Texte	5.113.249	2040+ Interviews und Tran- skriptionen	<ul style="list-style-type: none"> • <i>Habla Culta</i> • [ESP-ORAL] <i>Corpus oral de referencia de la lengua española contemporánea</i> • <i>Cortes de Castilla-La Mancha</i> • <i>Congreso de España</i> • <i>ABC Cultural 1991-1995</i> • <i>Chiapas (La República) – Interviews</i> • <i>Ernesto Zedillo – Interviews / Gespräche</i> • <i>Fidel Castro – Interviews / Gespräche</i> • <i>Partido PAN – Interviews</i> • <i>Partido PRI – Interviews</i>
Gesamt:	101.311.682	13.926	

Abbildung 134: Herkunft der Texte des CdE:Orig (Gesamt) nach Jahrhundert.¹³²⁸

¹³²⁸ Vgl. Davies: *Corpus del Español. Ayuda: Textos en el corpus*. Op. cit. (28.05.2011).

6. CdE:Orig – Einteilung der Texte des 20 Mio. Lexeme umfassenden Subkorpus des aktuellen Spanisch nach Genres und Registern

	Anzahl der Lexeme (in Mio.)	Spanien	Anzahl der Lexeme (in Mio.)	Lateinamerika
Gesprochen	1.00	<i>España Oral</i> ¹	2.00	<i>Habla Culta</i> (zehn Länder)
	0.35	<i>Habla Culta</i> (Madrid, Sevilla)		
3.35	1.35		2.00	
Transkripte/ Theaterstücke	1.00	Transkripte/ Interviews (Kongresse, Pressekonferenzen, andere)	1.00	Transkripte/ Interviews (Kongresse, Pressekonferenzen, andere)
	0.27	Interviews in der Zeitung		
	0.40	Theaterstücke	0.73	Theaterstücke
3.40	1.67		1.73	
Literatur	0.06	Romane (BV ²)	1.60	Romane (BV ²)
	0.00	Kurzgeschichten (BV ²)	0.87	Kurzgeschichten (BV ²)
	0.19	Drei Romane (BYU ³)	1.11	Zwölf Romane (BYU ³)
	2.17	Hauptsächlich Romane, aus LEXESP ⁴	0.18	Vier Romane aus Argentinien ⁵
			0.20	Drei Romane aus Chile ⁶
6.38	2.42		3.96	
Texte	1.05	Zeitung ABC	3.00	Zeitungen aus sechs verschiedenen Ländern
	0.15	Aufsätze in LEXESP ⁴	0.07	<i>Cartas</i> („Briefe“) aus Argentinien ⁵

8. Anhang

	2.00	<i>Encarta</i> Enzyklopädie	0.30	Humanistische Texte (z. B.: Philosophie, Geschichte Argentiniens ⁵)
			0.30	Humanistische Texte (z. B.: Philosophie, Geschichte Chiles ⁶)
6.87	3.20		3.67	
Gesamt	8.64		11.36	
<p>Genauere Angaben zu den oben aufgelisteten Quellen:</p> <p>¹ <i>Corpus oral de referencia de la lengua española contemporánea</i> (http://elvira.lllf.uam.es/docs/_es/corpus/corpus.html)</p> <p>² <i>Biblioteca Virtual</i> (http://www.cervantesvirtual.com)</p> <p>³ 15 aktuelle Romane in elektronischer Form, erworben vom <i>Humanities Research Center</i> der <i>Brigham Young University</i></p> <p>⁴ <i>Léxico informatizado del español</i> (http://www.edicionsub.com/colección.asp?colección=90)</p> <p>⁵ Aus dem <i>Corpus lingüístico de referencia de la lengua española en argentina</i> (http://www.lllf.uam.es/~fmarcos/informes/corpus/coarginl.html)</p> <p>⁶ Aus dem <i>Corpus lingüístico de referencia de la lengua española en chile</i> (http://www.lllf.uam.es/~fmarcos/informes/corpus/cochile.html)</p>				

Abbildung 135: Zusammensetzung des CdE:Orig nach Genres und Registern.¹³²⁹

¹³²⁹ Vgl. Davies (2005): op. cit., S. 107-108.

7. CORDE – Einteilung der Einträge nach Genre und Epoche

Genre	Mittelalter	<i>Siglos de Oro</i>	Zeitgenössische Epoche	Gesamt	Prozent
Vers-Lyrik	2.000.000	2.500.000	3.000.000	7.500.000	6%
Vers-Epik	3.250.000	1.000.000	750.000	5.000.000	4%
Vers-Drama	750.000	2.812.500	1.625.000	5.187.500	4,15%
Prosa-Lyrik	-	-	500.000	500.000	0,4%
Prosa-Narrativik	3.125.000	8.750.000	19.562.500	31.437.500	25,15%
Prosa-Drama	250.000	1.250.000	3.375.000	4.875.000	3,9%
Prosa-Didaktik	2.125.000	2.625.000	6.625.000	11.375.000	9,1%
Prosa-Wissenschaft	4.200.000	4.637.500	9.537.500	18.375.000	14,7%
Prosa-Gesellschaft	2.250.000	2.875.000	5.000.000	10.125.000	8,1%
Prosa-Presse	-	375.000	3.000.000	3.375.000	2,7%
Prosa-Religion	2.500.000	2.250.000	2.875.000	7.625.000	6,1%
Prosa-Geschichte	3.500.000	4.625.000	8.125.000	16.250.000	1,3%
Prosa-Rechtswesen	2.000.000	1.750.000	3.000.000	6.750.000	5,4%
Gesamt	25.950.000	35.075.000	63.975.000	125.000.000	100%

Abbildung 136: Anzahl der Einträge des CORDE in den verschiedenen Genres und Epochen.¹³³⁰

¹³³⁰ Die obige Tabelle wurde direkt von der Internetseite der *Real Academia Española* übernommen, lediglich die Angaben zu Genres und Epochen wurden ins Deutsche übertragen. Die auf der Internetseite angegebenen Zahlen lassen sich nicht zu den angegebenen Gesamtsummen addieren, wodurch die Übersicht rechnerisch inkorrekt ist. Dennoch wurde die Übersicht in die vorliegende Arbeit übernommen, da sie lediglich einen Überblick über die Herkunft der Texte geben soll, um die Textverteilung des CORDE zu veranschaulichen. Vgl. Real Academia Española (ed.): *CREA. Ayuda*. Op. cit. (14.05.2011).

8. Arbeitsblatt 1a: *Colocaciones* – Korpus: CORPES**Colocaciones**

Verwende das *Corpus del Español del Siglo XXI* (CORPES), um den anderen Teil der Kollokation zu ermitteln (<http://web.frl.es/CORPES/view/inicioExterno.view>). Zunächst wählst du die Funktion *Coapariciones* in der oberen Leiste aus, gibst die Basis der Kollokation ein (das unterstrichene Wort im jeweiligen Satz) und bestimmst die Wortart (z. B.: *sustantivo*). Nachdem du auf das Feld *Coapariciones* geklickt hast, wird eine Liste mit denjenigen Wörtern angezeigt, die vorzugsweise mit dieser Basis kombiniert werden. Nun suchst du ein passendes Wort der gesuchten Wortart vom Anfang der Liste aus, also mit einem hohen MI-Wert. Für den ersten Satz könnte somit das Verb *solucionar* ausgewählt werden. Verwende das Online-Wörterbuch PONS (<https://de.pons.com>), falls dir Wörter nicht bekannt sind.

Concordancias Coapariciones Configuración Ayuda Modo de cita Sugerencias	
Lema:	<input type="text" value="problema"/>
Clase de palabra:	<input type="text" value="sustantivo"/>
Origen	<div> (Todos) América España Filipinas </div>
<input type="button" value="Coapariciones"/> <input type="button" value="Nueva consulta"/>	

	Clase	Freq	MI	LL SIMPLE
solucionar	verbo	3254	9,83	17.143,12
acuciante	adjetivo	133	9,32	650,06
irresoluble	adjetivo	39	9,09	184,10
resolver	verbo	7080	8,98	33.072,04
aquejar	verbo	303	8,83	1.379,27
broncopulmonar	adjetivo	11	8,72	49,18
solventar	verbo	279	8,63	1.232,56

CORPES: Suche nach Kollokationen mit der Basis *problema*.

1. Esperamos (vb) este problema en el curso de la mañana para que todo esté listo esta tarde.
2. Desde la infancia nos unió una amistad (adj).
3. Cuando consiguieron salir, la lluvia (vb) y la calle estaba vacía.
4. Olivia había cometido un error (adj), sin duda, al enamorarse locamente de él y casarse pensando en ser feliz algún día.
5. Mucha gente lloró (adv) la muerte de Carolina.
6. Alba sonreía feliz ante la alegría (adj) del muchacho y el agradecimiento de la mirada del tío Pancho.
7. “Manolo, un momento, por favor: no quiero interrumpirte, pero te ruego (adv) que no uses términos despectivos: los negros, los moros, la selva...”.
8. A pesar de ser (adv) rico, Marcos vivía como un hombre solitario y pobre.

9. Quería comprar un libro para regalarlo a su novio Daniel, de quien estaba (adv) enamorada.
10. El policía preguntó: “¿Quién ha (vb) ese crimen tan despreciable?”
11. Muchos jóvenes orientan su vocación hacia el ámbito del arte y desean (adv) convertirse en artistas profesionales.
12. Casi puedo oler la cercanía del hombre. Y de la nada, sin yo sospecharlo, me (vb) un golpe en la cabeza. Pierdo el frágil equilibrio que me resta y caigo al suelo.
13. Sin embargo no pude (vb) el sueño como hubiera querido. Era un estado increíble de preocupación.
14. ¿Quién debía informarle al presidente sobre el accidente? – A esta pregunta siguió un silencio (adj).
15. Después del atentado presionaron al estado a (vb) medidas.
16. En apoyo de su teoría Manuel (vb) argumentos obvios: el dinero y el tiempo libre.
17. El estado tiene la obligación de (vb) estrategias para una solución pacífica.
18. ‘Siempre así’ es un grupo de diez jóvenes, cinco chicos y cinco chicas, amigos que un día decidieron tomarse en serio cantar y no paran de (vb) éxitos.
19. “No hay razón para alarma, no hay peligro (adj) para las poblaciones cercanas al Volcán Cerro Negro”, dijo la policía.
20. Me preocupa que Marisol no deje de vomitar, pero se niega (adv) a visitar a un médico.
21. El equipo no pudo lograr sus objetivos y fracasó (adv).
22. El más (adj) problema que debíamos resolver era el de integrar a Nathaniel a uno de los grupos.
23. El aire marino es de (adj) importancia terapéutica para estimular el sistema inmunitario.
24. Los medios de comunicación tendrán la oportunidad de (vb) preguntas al presidente.

Colocaciones – Solución

Verwende das *Corpus del Español del Siglo XXI* (CORPES), um den anderen Teil der Kollokation zu ermitteln (<http://web.frl.es/CORPES/view/inicioExterno.view>). Zunächst wählst du die Funktion *Coapariciones* in der oberen Leiste aus, gibst die Basis der Kollokation ein (das unterstrichene Wort im jeweiligen Satz) und bestimmst die Wortart (z. B.: *sustantivo*). Nachdem du auf das Feld *Coapariciones* geklickt hast, wird eine Liste mit denjenigen Wörtern angezeigt, die vorzugsweise mit dieser Basis kombiniert werden. Nun suchst du ein passendes Wort der gesuchten Wortart vom Anfang der Liste aus, also mit einem hohen MI-Wert. Für den ersten Satz könnte somit das Verb *solucionar* ausgewählt werden. Verwende das Online-Wörterbuch PONS (<https://de.pons.com>), falls dir Wörter nicht bekannt sind.

Concordancias | **Coapariciones** | Configuración | Ayuda | Modo de cita | Sugerencias

Lema: Clase de palabra:

Origen: (Todos) América España Filipinas

Coapariciones Nueva consulta

	Clase	Freq	MI	LL SIMPLE
solucionar	verbo	3254	9,83	17.143,12
acuciante	adjetivo	133	9,32	650,06
irresoluble	adjetivo	39	9,09	184,10
resolver	verbo	7080	8,98	33.072,04
aquejar	verbo	303	8,83	1.379,27
broncopulmonar	adjetivo	11	8,72	49,18
solventar	verbo	279	8,63	1.232,56

CORPES: Suche nach Kollokationen mit der Basis *problema*.

1. Esperamos [solucionar] este problema en el curso de la mañana para que todo esté listo esta tarde.
2. Desde la infancia nos unió una amistad [entrañable].
3. Cuando consiguieron salir, la lluvia [arreciaba] y la calle estaba vacía.
4. Olivia había cometido un error [garrafal], sin duda, al enamorarse locamente de él y casarse pensando en ser feliz algún día.
5. Mucha gente lloró [desconsoladamente] la muerte de Carolina.
6. Alba sonreía feliz ante la alegría [desbordante] del muchacho y el agradecimiento de la mirada del tío Pancho.
7. “Manolo, un momento, por favor: no quiero interrumpirte, pero te ruego [encarecidamente] que no uses términos despectivos: los negros, los moros, la selva...”
8. A pesar de ser [inmensamente] rico, Marcos vivía como un hombre solitario y pobre.
9. Quería comprar un libro para regalarlo a su novio Daniel, de quien estaba [perdidamente] enamorado.
10. El policía preguntó: “¿Quién ha [cometido] ese crimen tan despreciable?”
11. Muchos jóvenes orientan su vocación hacia el ámbito del arte y desean [fervientemente] convertirse en artistas profesionales.

12. Casi puedo oler la cercanía del hombre. Y de la nada, sin yo sospecharlo, me [asesta] un golpe en la cabeza. Pierdo el frágil equilibrio que me resta y caigo al suelo.
13. Sin embargo no pude [conciliar] el sueño como hubiera querido. Era un estado increíble de preocupación.
14. ¿Quién debía informarle al presidente sobre el accidente? – A esta pregunta siguió un silencio [sepulcral].
15. Después del atentado presionaron al estado a [adoptar] medidas.
16. En apoyo de su teoría Manuel [esgrime] argumentos obvios: el dinero y el tiempo libre.
17. El estado tiene la obligación de [implementar] estrategias para una solución pacífica.
18. ‘Siempre así’ es un grupo de diez jóvenes, cinco chicos y cinco chicas, amigos que un día decidieron tomarse en serio cantar y no paran de [cosechar] éxitos.
19. “No hay razón para alarma, no hay peligro [inminente] para las poblaciones cercanas al Volcán Cerro Negro”, dijo la policía.
20. Me preocupa que Marisol no deje de vomitar, pero se niega [rotundamente] a visitar a un médico.
21. El equipo no pudo lograr sus objetivos y fracasó [estrepitosamente].
22. El más [acuciante] problema que debíamos resolver era el de integrar a Nathaniel a uno de los grupos.
23. El aire marino es de [suma] importancia terapéutica para estimular el sistema inmunitario.
24. Los medios de comunicación tendrán la oportunidad de [formular] preguntas al presidente.

9. Arbeitsblatt 1b: *Colocaciones* – Korpus: CORPES**Colocaciones**

Verwende das *Corpus del Español del Siglo XXI* (CORPES), um den anderen Teil der Kollokation zu ermitteln (<http://web.frl.es/CORPES/view/inicioExterno.view>). Zunächst wählst du die Funktion *Coapariciones* in der oberen Leiste aus, gibst die Basis der Kollokation ein (das unterstrichene Wort im jeweiligen Satz) und bestimmst die Wortart (z. B.: *sustantivo*). Nachdem du auf das Feld *Coapariciones* geklickt hast, wird eine Liste mit denjenigen Wörtern angezeigt, die vorzugsweise mit dieser Basis kombiniert werden. Nun suchst du ein passendes Wort der gesuchten Wortart vom Anfang der Liste aus, also mit einem hohen MI-Wert. Für den ersten Satz könnte somit das Verb *solucionar* ausgewählt werden. Verwende das Online-Wörterbuch PONS (<https://de.pons.com>), falls dir Wörter nicht bekannt sind.

Concordancias | **Coapariciones** | Configuración | Ayuda | Modo de cita | Sugerencias

Lema: Clase de palabra:

Origen:

Coapariciones Nueva consulta

	Clase	Freq	MI	LL SIMPLE
solucionar	verbo	3254	9,83	17.143,12
acuciante	adjetivo	133	9,32	650,06
irresoluble	adjetivo	39	9,09	184,10
resolver	verbo	7080	8,98	33.072,04
aquejar	verbo	303	8,83	1.379,27
broncopulmonar	adjetivo	11	8,72	49,18
solventar	verbo	279	8,63	1.232,56

CORPES: Suche nach Kollokationen mit der Basis *problema*.

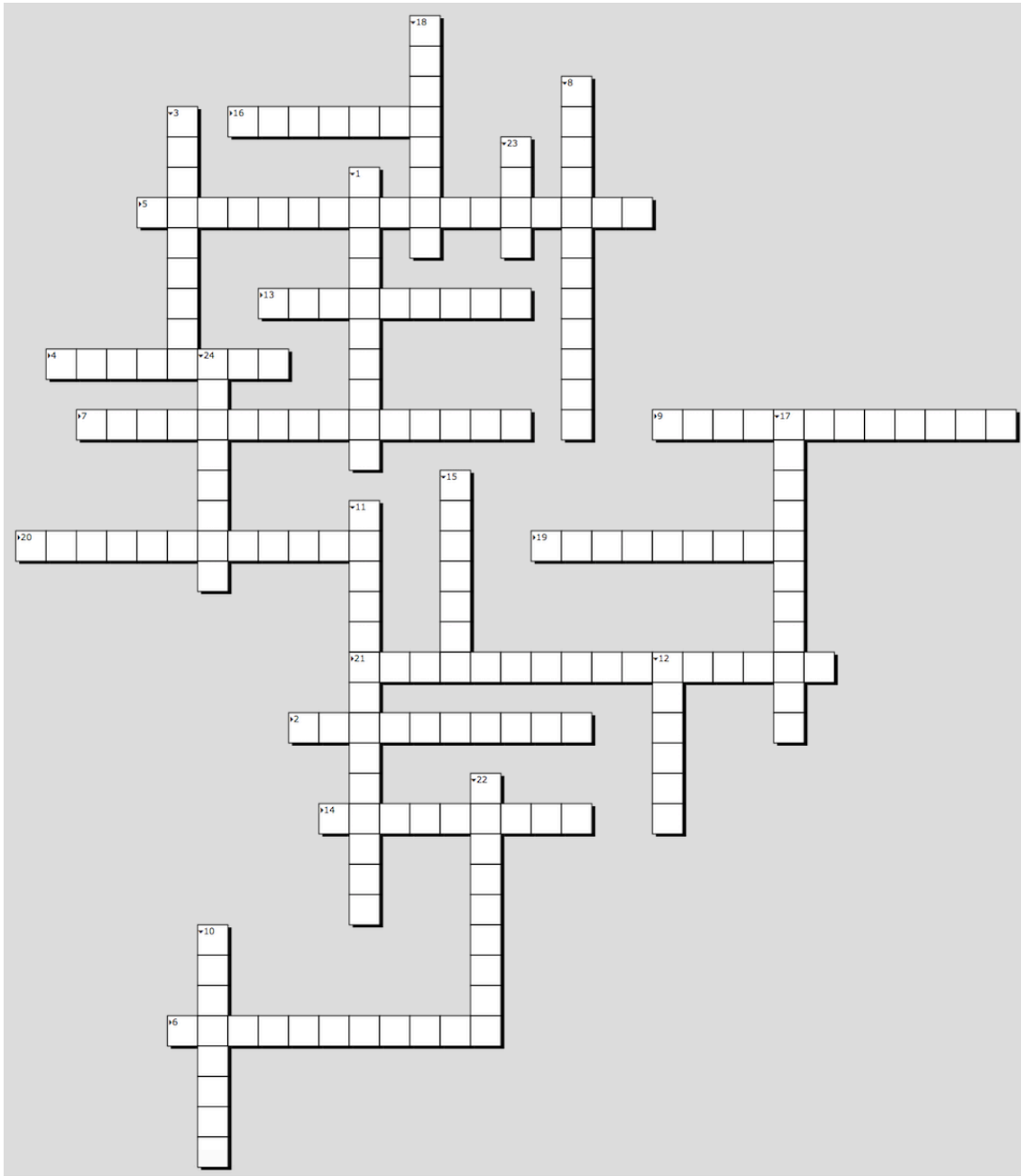
1. Esperamos (vb) este problema en el curso de la mañana para que todo esté listo esta tarde.
2. Desde la infancia nos unió una amistad (adj).
3. Cuando consiguieron salir, la lluvia (vb) y la calle estaba vacía.
4. Olivia había cometido un error (adj), sin duda, al enamorarse locamente de él y casarse pensando en ser feliz algún día.
5. Mucha gente lloró (adv) la muerte de Carolina.
6. Alba sonreía feliz ante la alegría (adj) del muchacho y el agradecimiento de la mirada del tío Pancho.
7. “Manolo, un momento, por favor: no quiero interrumpirte, pero te ruego (adv) que no uses términos despectivos: los negros, los moros, la selva...”.
8. A pesar de ser (adv) rico, Marcos vivía como un hombre solitario y pobre.
9. Quería comprar un libro para regalarlo a su novio Daniel, de quien estaba (adv) enamorada.

10. El policía preguntó: “¿Quién ha (vb) ese crimen tan despreciable?”
11. Muchos jóvenes orientan su vocación hacia el ámbito del arte y desean (adv) convertirse en artistas profesionales.
12. Casi puedo oler la cercanía del hombre. Y de la nada, sin yo sospecharlo, me (vb) un golpe en la cabeza. Pierdo el frágil equilibrio que me resta y caigo al suelo.
13. Sin embargo no pude (vb) el sueño como hubiera querido. Era un estado increíble de preocupación.
14. ¿Quién debía informarle al presidente sobre el accidente? – A esta pregunta siguió un silencio (adj).
15. Después del atentado presionaron al estado a (vb) medidas.
16. En apoyo de su teoría Manuel (vb) argumentos obvios: el dinero y el tiempo libre.
17. El estado tiene la obligación de (vb) estrategias para una solución pacífica.
18. ‘Siempre así’ es un grupo de diez jóvenes, cinco chicos y cinco chicas, amigos que un día decidieron tomarse en serio cantar y no paran de (vb) éxitos.
19. “No hay razón para alarma, no hay peligro (adj) para las poblaciones cercanas al Volcán Cerro Negro”, dijo la policía.
20. Me preocupa que Marisol no deje de vomitar, pero se niega (adv) a visitar a un médico.
21. El equipo no pudo lograr sus objetivos y fracasó (adv).
22. El más (adj) problema que debíamos resolver era el de integrar a Nathaniel a uno de los grupos.
23. El aire marino es de (adj) importancia terapéutica para estimular el sistema inmunitario.
24. Los medios de comunicación tendrán la oportunidad de (vb) preguntas al presidente.

Después de haber solucionado este ejercicio introduce las palabras en el crucigrama¹³³¹ siguiente:

¹³³¹ Erstellung des Kreuzworträtsels mit dem im Internet kostenlos nutzbaren Programm *XWords*. Vgl. Ebert (2012-2017): op. cit. (26.10.2017).

8. Anhang



Colocaciones – Solución

Verwende das *Corpus del Español del Siglo XXI* (CORPES), um den anderen Teil der Kollokation zu ermitteln (<http://web.frl.es/CORPES/view/inicioExterno.view>). Zunächst wählst du die Funktion *Coapariciones* in der oberen Leiste aus, gibst die Basis der Kollokation ein (das unterstrichene Wort im jeweiligen Satz) und bestimmst die Wortart (z. B.: *sustantivo*). Nachdem du auf das Feld *Coapariciones* geklickt hast, wird eine Liste mit denjenigen Wörtern angezeigt, die vorzugsweise mit dieser Basis kombiniert werden. Nun suchst du ein passendes Wort der gesuchten Wortart vom Anfang der Liste aus, also mit einem hohen MI-Wert. Für den ersten Satz könnte somit das Verb *solucionar* ausgewählt werden. Verwende das Online-Wörterbuch PONS (<https://de.pons.com>), falls dir Wörter nicht bekannt sind.

Concordancias | **Coapariciones** | Configuración | Ayuda | Modo de cita | Sugerencias

Lema: Clase de palabra:

Origen:

Coapariciones Nueva consulta

	Clase	Freq	MI	LL SIMPLE
solucionar	verbo	3254	9,83	17.143,12
acuciante	adjetivo	133	9,32	650,06
irresoluble	adjetivo	39	9,09	184,10
resolver	verbo	7080	8,98	33.072,04
aquejar	verbo	303	8,83	1.379,27
broncopulmonar	adjetivo	11	8,72	49,18
solventar	verbo	279	8,63	1.232,56

CORPES: Suche nach Kollokationen mit der Basis *problema*.

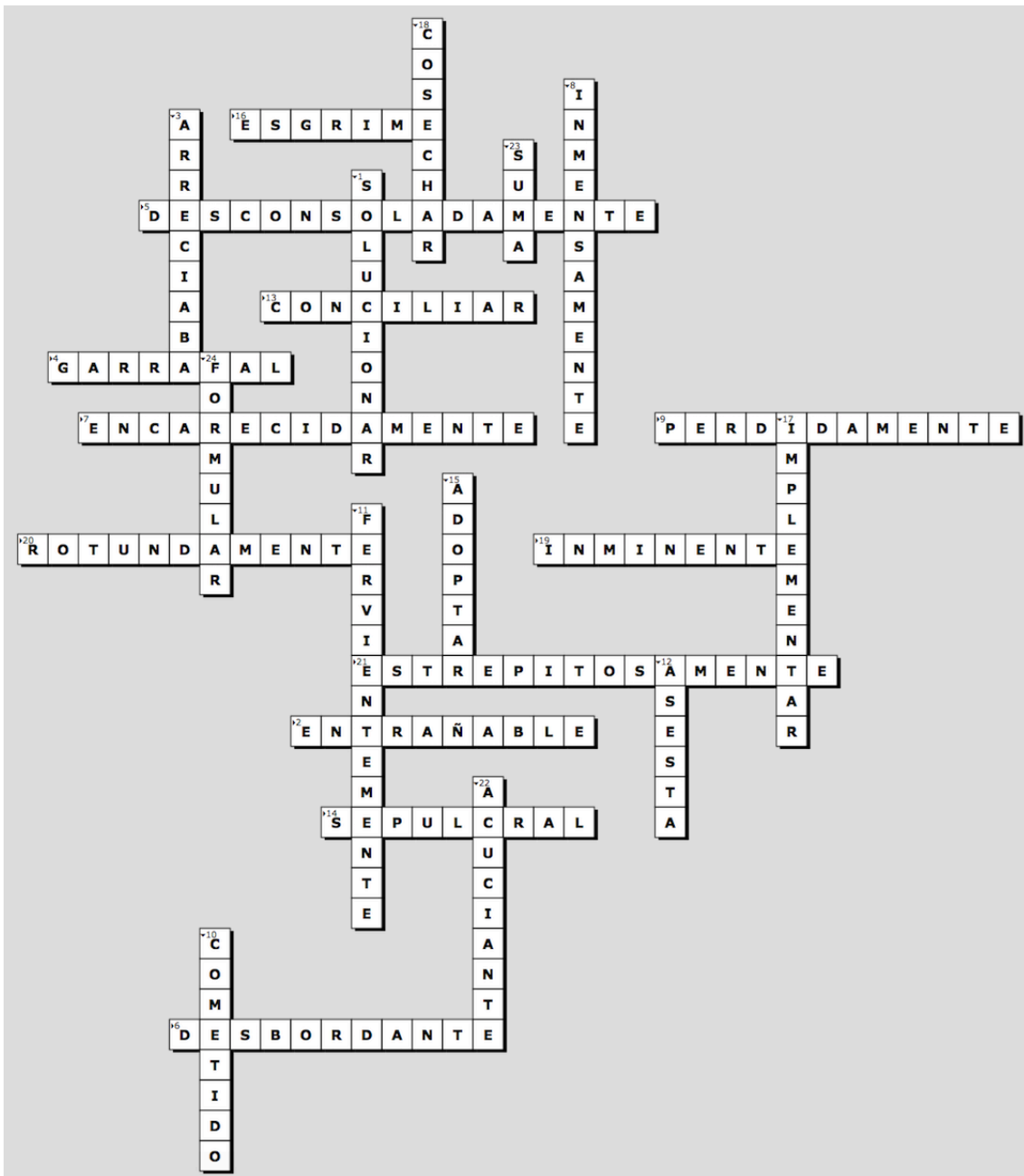
1. Esperamos [solucionar] este problema en el curso de la mañana para que todo esté listo esta tarde.
2. Desde la infancia nos unió una amistad [entrañable].
3. Cuando consiguieron salir, la lluvia [arreciaba] y la calle estaba vacía.
4. Olivia había cometido un error [garrafal], sin duda, al enamorarse locamente de él y casarse pensando en ser feliz algún día.
5. Mucha gente lloró [desconsoladamente] la muerte de Carolina.
6. Alba sonreía feliz ante la alegría [desbordante] del muchacho y el agradecimiento de la mirada del tío Pancho.
7. “Manolo, un momento, por favor: no quiero interrumpirte, pero te ruego [encarecidamente] que no uses términos despectivos: los negros, los moros, la selva...”
8. A pesar de ser [inmensamente] rico, Marcos vivía como un hombre solitario y pobre.
9. Quería comprar un libro para regalarlo a su novio Daniel, de quien estaba [perdidamente] enamorada.
10. El policía preguntó: “¿Quién ha [cometido] ese crimen tan despreciable?”

8. Anhang

11. Muchos jóvenes orientan su vocación hacia el ámbito del arte y desean [fervientemente] convertirse en artistas profesionales.
12. Casi puedo oler la cercanía del hombre. Y de la nada, sin yo sospecharlo, me [asesta] un golpe en la cabeza. Pierdo el frágil equilibrio que me resta y caigo al suelo.
13. Sin embargo no pude [conciliar] el sueño como hubiera querido. Era un estado increíble de preocupación.
14. ¿Quién debía informarle al presidente sobre el accidente? – A esta pregunta siguió un silencio [sepulcral].
15. Después del atentado presionaron al estado a [adoptar] medidas.
16. En apoyo de su teoría Manuel [esgrime] argumentos obvios: el dinero y el tiempo libre.
17. El estado tiene la obligación de [implementar] estrategias para una solución pacífica.
18. ‘Siempre así’ es un grupo de diez jóvenes, cinco chicos y cinco chicas, amigos que un día decidieron tomarse en serio cantar y no paran de [cosechar] éxitos.
19. “No hay razón para alarma, no hay peligro [inminente] para las poblaciones cercanas al Volcán Cerro Negro”, dijo la policía.
20. Me preocupa que Marisol no deje de vomitar, pero se niega [rotundamente] a visitar a un médico.
21. El equipo no pudo lograr sus objetivos y fracasó [estrepitosamente].
22. El más [acuciante] problema que debíamos resolver era el de integrar a Nathaniel a uno de los grupos.
23. El aire marino es de [suma] importancia terapéutica para estimular el sistema inmunitario.
24. Los medios de comunicación tendrán la oportunidad de [formular] preguntas al presidente.

Después de haber solucionado este ejercicio introduce las palabras en el crucigrama¹³³² siguiente:

¹³³² Erstellung des Kreuzworträtsels mit dem im Internet kostenlos nutzbaren Programm *XWords*. Vgl. Ebert (2012-2017): op. cit. (26.10.2017).



10. Arbeitsblatt 2a: *En busca de colocaciones* – Korpus: CORPES**En busca de colocaciones**

Verwende das *Corpus del Español del Siglo XXI* (CORPES), um den anderen Teil der Kollokation zu ermitteln (<http://web.frl.es/CORPES/view/inicioExterno.view>). Zunächst wählst du die Funktion *Coapariciones* in der oberen Leiste aus, gibst die Basis der Kollokation ein (das unterstrichene Wort im jeweiligen Satz) und bestimmst die Wortart (z. B.: *sustantivo*). Nachdem du auf das Feld *Coapariciones* geklickt hast, wird eine Liste mit denjenigen Wörtern angezeigt, die vorzugsweise mit dieser Basis kombiniert werden. Nun suchst du ein passendes Wort der gesuchten Wortart vom Anfang der Liste aus, also mit einem hohen MI-Wert. Für den ersten Satz könnte somit das Verb *solucionar* ausgewählt werden, welches du in die erste Spalte einträgst. Gib die Übersetzung der Kollokation in der zweiten Spalte an. Schließlich vervollständigst du den angegebenen Satz mit dem fehlenden Teil der Kollokation in der richtigen Form. Verwende das Online-Wörterbuch PONS (<https://de.pons.com>), falls dir Wörter nicht bekannt sind.

Concordancias | **Coapariciones** | Configuración | Ayuda | Modo de cita | Sugerencias

Lema: Clase de palabra:

Origen:

Coapariciones Nueva consulta

	Clase	Freq	MI	LL SIMPLE
solucionar	verbo	3254	9,83	17.143,12
acuciante	adjetivo	133	9,32	650,06
irresoluble	adjetivo	39	9,09	184,10
resolver	verbo	7080	8,98	33.072,04
aquejar	verbo	303	8,83	1.379,27
broncopulmonar	adjetivo	11	8,72	49,18
solventar	verbo	279	8,63	1.232,56

CORPES: Suche nach Kollokationen mit der Basis *problema*.

Colocación	Significado alemán	Ejemplo
un problema		Tengo que (vb) este <u>problema</u> con el jefe.
error		He cometido un <u>error</u> (adj) en el examen esta mañana.
llorar		<u>Lloré</u> (adv) después de la muerte de mi perro.
desear		<u>Deseo</u> (adv) convertirme en una actriz famosa.

_____		Juan me (vb)
un golpe		un <u>golpe</u> en la cabeza.
_____		No pude (vb)
el sueño		el <u>sueño</u> la noche pasada, porque tenía que pensar en el examen.
rogar		"Carla, te <u>ruego</u>
_____	 (adv) que no molestes a tu hermana."
guardar		Es un secreto que <u>guardamos</u>
_____	 (adv).
rechazar		La propuesta <u>fue rechazada</u>
_____	 (adv).
lluvia		Llegué tarde debido a las <u>lluvias</u>
_____	 (adj).
_____		Se ha hecho
rico		(adj) <u>rico</u> por sus negocios inmobiliarios.
_____		Pablo (vb) un
un crimen		<u>crimen</u> .
_____		En Buenos Aires
amistad	 (vb) <u>amistad</u> con Julia.
_____		Su opinión es
opuesto		(adv) <u>opuesta</u> a la mía.
_____		Estoy (adv)
enamorado		<u>enamorado</u> de Miguel.
_____		Resolver este problema es
imposible	 (adv)
_____		<u>imposible</u> .
fracasar		<u>Fracasaban</u>
_____		(adv) cuando intentaban descifrar este enigma.

En busca de colocaciones – *Solución*

Verwende das *Corpus del Español del Siglo XXI* (CORPES), um den anderen Teil der Kollokation zu ermitteln (<http://web.frl.es/CORPES/view/inicioExterno.view>). Zunächst wählst du die Funktion *Coapariciones* in der oberen Leiste aus, gibst die Basis der Kollokation ein (das unterstrichene Wort im jeweiligen Satz) und bestimmst die Wortart (z. B.: *sustantivo*). Nachdem du auf das Feld *Coapariciones* geklickt hast, wird eine Liste mit denjenigen Wörtern angezeigt, die vorzugsweise mit dieser Basis kombiniert werden. Nun suchst du ein passendes Wort der gesuchten Wortart vom Anfang der Liste aus, also mit einem hohen MI-Wert. Für den ersten Satz könnte somit das Verb *solucionar* ausgewählt werden, welches du in die erste Spalte einträgst. Gib die Übersetzung der Kollokation in der zweiten Spalte an. Schließlich vervollständigst du den angegebenen Satz mit dem fehlenden Teil der Kollokation in der richtigen Form. Verwende das Online-Wörterbuch PONS (<https://de.pons.com>), falls dir Wörter nicht bekannt sind.

Concordancias Coapariciones Configuración Ayuda Modo de cita Sugerencias	
Lema:	<input type="text" value="problema"/>
Clase de palabra:	<input type="text" value="sustantivo"/>
Origen	<div> (Todos) América España Filipinas </div>
<div> Coapariciones Nueva consulta </div>	

	Clase	Freq	MI	LL SIMPLE
<u>solucionar</u>	verbo	3254	9,83	17.143,12
<u>acuciante</u>	adjetivo	133	9,32	650,06
<u>irresoluble</u>	adjetivo	39	9,09	184,10
<u>resolver</u>	verbo	7080	8,98	33.072,04
<u>aquejar</u>	verbo	303	8,83	1.379,27
<u>broncopulmonar</u>	adjetivo	11	8,72	49,18
<u>solventar</u>	verbo	279	8,63	1.232,56

CORPES: Suche nach Kollokationen mit der Basis *problema*.

Colocación	Significado alemán	Ejemplo
<u>solucionar/ resolver</u> un problema	ein Problem lösen	Tengo que <u>solucionar</u> este <u>problema</u> con el jefe.
error <u>garrafal</u>	ein grober Fehler	He cometido un <u>error garrafal</u> en el examen esta mañana.
llorar <u>desconsoladamente</u>	untröstlich weinen	<u>Lloré desconsoladamente</u> después de la muerte de mi perro.
desear <u>fervientemente</u>	zutiefst wünschen	<u>Deseo fervientemente</u> convertirme en una actriz famosa.
<u>asestar</u> un golpe	einen Schlag versetzen	Juan me <u>asestó</u> un <u>golpe</u> en la cabeza.

<u>conciliar</u> el sueño	einschlafen	No pude <u>conciliar</u> el <u>sueño</u> la noche pasada, porque tenía que pensar en el examen.
rogar <u>encarecidamente</u>	eindringlich bitten	“Carla, te <u>ruego encarecidamente</u> que no molestes a tu hermana.”
guardar <u>celosamente</u>	sorgsam behüten, wie den eigenen Augapfel behüten	Es un secreto que <u>guardamos celosamente</u> .
rechazar <u>de plano</u>	weit von sich weisen/ kategorisch ablehnen	La propuesta <u>fue rechazada de plano</u> .
lluvia <u>torrencial</u>	sintflutartiger Regen	Llegué tarde debido a las <u>lluvias torrenciales</u> .
<u>inmensamente</u> rico	sehr reich, steinreich	Se ha hecho <u>inmensamente rico</u> por sus negocios inmobiliarios.
<u>cometer</u> un crimen	ein Verbrechen begehen	Pablo <u>cometió un crimen</u> .
<u>trabar/ entablar</u> amistad	Freundschaft schließen/ sich befreunden	En Buenos Aires <u>trabé amistad</u> con Julia.
<u>diametralmente</u> opuesto	genau entgegengesetzt	Su opinión es <u>diametralmente opuesta</u> a la mía.
<u>perdidamente/ locamente</u> enamorado	unsterblich verliebt	Estoy <u>perdidamente enamorada</u> de Miguel.
<u>de todo punto/ materialmente/ prácticamente</u> imposible	absolut unmöglich/ praktisch unmöglich	Resolver este problema es <u>de todo punto imposible</u> .
fracasar <u>estrepitosamente</u>	jämmerlich scheitern	<u>Fracasaban estrepitosamente</u> cuando intentaban descifrar este enigma.

11. Arbeitsblatt 2b: *En busca de colocaciones* – Korpus: CORPES**En busca de colocaciones**

Verwende das *Corpus del Español del Siglo XXI* (CORPES), um den anderen Teil der Kollokation zu ermitteln (<http://web.frl.es/CORPES/view/inicioExterno.view>). Zunächst wählst du die Funktion *Coapariciones* in der oberen Leiste aus, gibst die Basis der Kollokation ein (das unterstrichene Wort im jeweiligen Satz) und bestimmst die Wortart (z. B.: *sustantivo*). Nachdem du auf das Feld *Coapariciones* geklickt hast, wird eine Liste mit denjenigen Wörtern angezeigt, die vorzugsweise mit dieser Basis kombiniert werden. Nun suchst du ein passendes Wort der gesuchten Wortart vom Anfang der Liste aus, also mit einem hohen MI-Wert. Für den ersten Satz könnte somit das Verb *solucionar* ausgewählt werden, welches du in die erste Spalte einträgst. Schließlich verfasst du einen Satz mit der jeweiligen Kollokation – für diese Aufgabe kannst du die Beispiele im CORPES nutzen. Verwende das Online-Wörterbuch PONS (<https://de.pons.com>), falls dir Wörter nicht bekannt sind.

Concordancias | **Coapariciones** | Configuración | Ayuda | Modo de cita | Sugerencias

Lema: Clase de palabra:

Origen:

Coapariciones Nueva consulta

	Clase	Freq	MI	LL SIMPLE
solucionar	verbo	3254	9,83	17.143,12
acuciante	adjetivo	133	9,32	650,06
irresoluble	adjetivo	39	9,09	184,10
resolver	verbo	7080	8,98	33.072,04
aquejar	verbo	303	8,83	1.379,27
broncopulmonar	adjetivo	11	8,72	49,18
solventar	verbo	279	8,63	1.232,56

CORPES: Suche nach Kollokationen mit der Basis *problema*.

Colocación	Ejemplo
1. _____ un problema (vb)	Tengo que <u>solucionar</u> este <u>problema</u> con el jefe.
2. error _____ (adj)	
3. llorar _____ (adv)	
4. desear _____ (adv)	

5. _____ un golpe (vb)	
6. _____ el sueño (vb)	
7. rogar _____ (adv)	
8. _____ imposible (adv)	
9. rechazar _____ (adv)	
10. lluvia _____ (adj)	
11. _____ rico (adv)	
12. _____ un crimen (vb)	
13. fracasar _____ (adv)	
14. _____ opuesto (adv)	
15. _____ enamorado (adv)	

En busca de colocaciones – Solución

Verwende das *Corpus del Español del Siglo XXI* (CORPES), um den anderen Teil der Kollokation zu ermitteln (<http://web.frl.es/CORPES/view/inicioExterno.view>). Zunächst wählst du die Funktion *Coapariciones* in der oberen Leiste aus, gibst die Basis der Kollokation ein (das unterstrichene Wort im jeweiligen Satz) und bestimmst die Wortart (z. B.: *sustantivo*). Nachdem du auf das Feld *Coapariciones* geklickt hast, wird eine Liste mit denjenigen Wörtern angezeigt, die vorzugsweise mit dieser Basis kombiniert werden. Nun suchst du ein passendes Wort der gesuchten Wortart vom Anfang der Liste aus, also mit einem hohen MI-Wert. Für den ersten Satz könnte somit das Verb *solucionar* ausgewählt werden, welches du in die erste Spalte einträgst. Schließlich verfasst du einen Satz mit der jeweiligen Kollokation – für diese Aufgabe kannst du die Beispiele im CORPES nutzen. Verwende das Online-Wörterbuch PONS (<https://de.pons.com>), falls dir Wörter nicht bekannt sind.

Concordancias Coapariciones Configuración Ayuda Modo de cita Sugerencias	
Lema:	<input type="text" value="problema"/>
Clase de palabra:	<input type="text" value="sustantivo"/>
Origen	<input type="text" value="(Todos)"/> <input type="text" value="América"/> <input type="text" value="España"/> <input type="text" value="Filipinas"/>
<input type="button" value="Coapariciones"/> <input type="button" value="Nueva consulta"/>	

	Clase	Freq	MI	LL SIMPLE
solucionar	verbo	3254	9,83	17.143,12
acuciante	adjetivo	133	9,32	650,06
irresoluble	adjetivo	39	9,09	184,10
resolver	verbo	7080	8,98	33.072,04
aquejar	verbo	303	8,83	1.379,27
broncopulmonar	adjetivo	11	8,72	49,18
solventar	verbo	279	8,63	1.232,56

CORPES: Suche nach Kollokationen mit der Basis *problema*.

Colocación	Ejemplo
1. <u>solucionar/ resolver</u> un problema	<i>Tengo que <u>solucionar</u> este <u>problema</u> con el jefe.</i>
2. error <u>garrafal</u>	<i>He cometido un <u>error garrafal</u> en el examen esta mañana.</i>
3. llorar <u>desconsoladamente</u>	<i><u>Lloré desconsoladamente</u> después de la muerte de mi perro.</i>
4. desear <u>fervientemente</u>	<i><u>Deseo fervientemente</u> convertirme en una actriz famosa.</i>
5. <u>asestar</u> un golpe	<i>Juan me <u>asestó</u> un <u>golpe</u> en la cabeza.</i>
6. <u>conciliar</u> el sueño	<i>No pude <u>conciliar</u> el <u>sueño</u> la noche pasada, porque tenía que pensar en el examen.</i>

7. rogar <u>encarecidamente</u>	<i>"Carla, te <u>ruego encarecidamente</u> que no molestes a tu hermana."</i>
8. <u>de todo punto/</u> <u>materialmente/</u> <u>prácticamente</u> imposible	<i>Resolver este problema es <u>prácticamente imposible</u>.</i>
9. rechazar <u>de plano</u>	<i>La propuesta <u>fue rechazada de plano</u>.</i>
10. lluvia <u>torrencial</u>	<i>Llegué tarde debido a las <u>lluvias torrenciales</u>.</i>
11. <u>inmensamente</u> rico	<i>Se ha hecho <u>inmensamente rico</u> por sus negocios inmobiliarios.</i>
12. <u>cometer</u> un crimen	<i>Hace veinte años Pablo <u>cometió un crimen</u>.</i>
13. fracasar <u>estrepitosamente</u>	<i><u>Fracasaban estrepitosamente</u> cuando intentaban descifrar este enigma.</i>
14. <u>diametralmente</u> opuesto	<i>Su opinión es <u>diametralmente opuesta</u> a la mía.</i>
15. <u>perdidamente/</u> <u>locamente</u> enamorado	<i>Estoy <u>perdidamente enamorada</u> de Miguel.</i>

12. Arbeitsblatt 3: *Mapa mental: Aspecto* – Korpus: CORPES**Mapa mental: Aspecto**

Verwende das *Corpus del Español del Siglo XXI* (CORPES), um Vokabular für die Erstellung einer Mindmap zum Thema „Aussehen“ zu finden (http://web.frl.es/CORPES/view/inicio_Externo.view). Zunächst wählst du die Funktion *Coapariciones* in der oberen Leiste aus, gibst den Suchbegriff ein (z. B.: *pelo*), bestimmst die Wortart (z. B.: *sustantivo*) und klickst auf das Feld *Coapariciones*. Daraufhin klickst du zweimal auf die Angabe *LL-Simple* und einmal auf *Clase*, um die Treffer zunächst in absteigender Reihenfolge nach der statistischen Methode *LL-Simple* und schließlich nach der Wortart zu sortieren. Aus der erstellten Liste, welche Wörter enthält, die zusammen mit dem Suchbegriff verwendet werden (angefangen bei Adjektiven bis hin zu Verben), wählst du einige Begriffe aus, um die Mindmap auszufüllen. Wenn du auf einen der blau geschriebenen Treffer klickst, erscheinen Beispielsätze mit der jeweiligen Kombination (z. B.: *pelo largo*). Verwende das Online-Wörterbuch PONS (<https://de.pons.com>), falls dir Wörter nicht bekannt sind.

Concordancias | **Coapariciones** | Configuración | Ayuda | Modo de cita | Sugerencias

Lema: Clase de palabra: Tema:

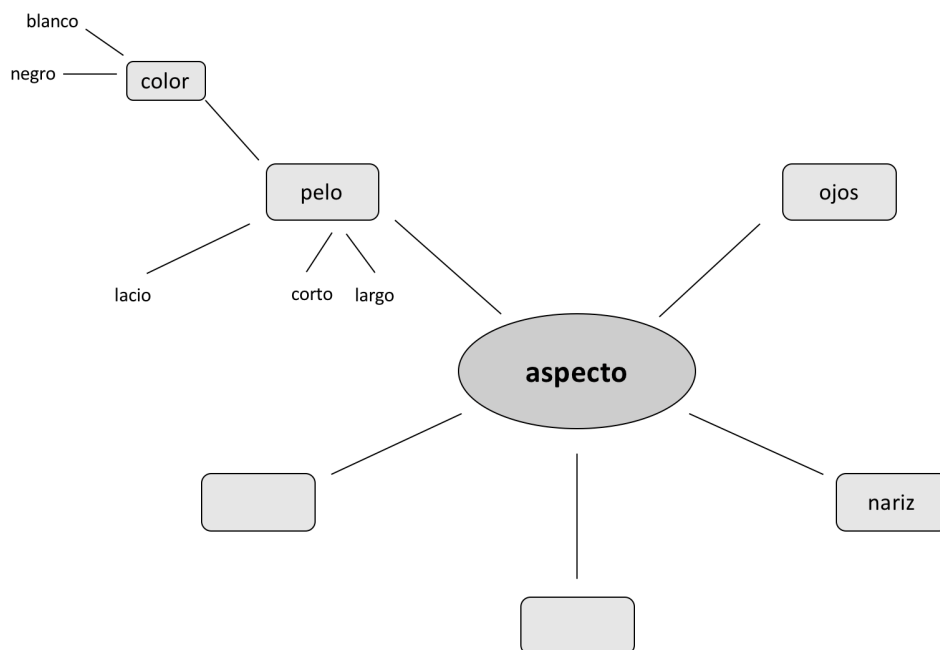
Origen:

Coapariciones Nueva consulta

	Clase	Freq	MI	LL SIMPLE	T-SCORE
largo	adjetivo	1377	7,98	5.474,91	36,97
negro	adjetivo	1163	8,14	4.738,73	33,98
corto	adjetivo	759	8,59	3.294,49	27,51
blanco	adjetivo	792	7,58	2.952,00	28,00
rubio	adjetivo	531	10,44	2.911,25	23,04

CORPES: Suche nach Wörtern, die mit dem Substantiv *pelo* zusammen verwendet werden.

Copia el mapa mental siguiente en tu cuaderno de ejercicios y rellénalo con palabras apropiadas:



13. Arbeitsblatt 4: *Adjetivos adecuados* – Korpus: CdE:Orig (Version 2008)**Adjetivos adecuados****¿Qué adjetivos se usan con estos sustantivos?**

Verwende die spanische Version des *Corpus del Español: Género/Histórico* mit dem alten Interface aus dem Jahr 2008, um Adjektive zu finden, die zusammen mit den unten angegebenen Substantiven verwendet werden (<https://www.corpusdelespanol.org/hist-gen/2008/x.asp>). Zunächst gibst du das Substantiv in eckigen Klammern in die Zeile *Palabra(s)* ein. Daraufhin klickst du auf das Fragezeichen in der Zeile *Colocados* und wählst eine Kollokationsspanne von je einem Wort auf der linken und rechten Seite aus. Schließlich klickst du auf das Fragezeichen in der Zeile *Cat Gram* und bestimmst *adj.Todos* als Wortart. Auf diese

Weise kannst du nach Adjektiven suchen, die unmittelbar neben dem jeweiligen Substantiv stehen. Nun klickst du auf *Buscar*, schreibst einige passende Adjektive aus der angezeigten Liste in die zweite und einen Beispielsatz in die dritte Spalte der Tabelle. Für den Beispielsatz kannst du Sätze aus dem Korpus verwenden, zu denen du gelangst, indem du auf die blau geschriebenen Treffer klickst. Nutze das Online-Wörterbuch PONS (<https://de.pons.com>), falls dir Wörter nicht bekannt sind.

MOstrar

• LISTA • GRÁFICO • PCEC • COMPARAR

BUSCAR

PALABRA(S) [ventaja]

COLOCADOS [J*] 1 1

CAT GRAM adj.TODOS

AL AZAR BUSCAR BORRAR

SECCIONES MOstrar

1 -IGNORAR- s20 s19 s18 s17 s16 s15

2 -IGNORAR- s20 s19 s18 s17 s16 s15

ORDENAR Y LIMITAR

ORDENAR FRECUENCIA

MÍNIMO FRECUENCIA 4

Interface des CdE:Orig (Version 2008).

Sustantivo	Adjetivos adecuados	Ejemplo
ventaja	gran(de), mayor, cierta, inmensa, principal, notable, enorme, clara, importante, única, doble, amplia, adicional, aparente, considerable	Estas máquinas presentan <u>ventajas considerables</u> sobre la fabricación manual.
problema		
objetivo		
hecho		

8. Anhang

precio		
argumento		
asunto		
razón		
cuestión		
error		
amigo		
idea		
plan		

Adjetivos adecuados -

Solución

¿Qué adjetivos se usan con estos sustantivos?

Verwende die spanische Version des *Corpus del Español: Género/Histórico* mit dem alten Interface aus dem Jahr 2008, um Adjektive zu finden, die zusammen mit den unten angegebenen Substantiven verwendet werden (<https://www.corpusdelespanol.org/hist-gen/2008/x.asp>). Zunächst gibst du das Substantiv in eckigen Klammern in die Zeile *Palabra(s)* ein. Daraufhin klickst du auf das Fragezeichen in der Zeile *Colocados* und wählst eine Kollokationsspanne von je einem Wort auf der linken und rechten Seite aus. Schließlich klickst du auf das Fragezeichen in der Zeile *Cat Gram* und

Interface des CdE:Orig (Version 2008).

bestimmst *adj.Todos* als Wortart. Auf diese Weise kannst du nach Adjektiven suchen, die unmittelbar neben dem jeweiligen Substantiv stehen. Nun klickst du auf *Buscar*, schreibst einige passende Adjektive aus der angezeigten Liste in die zweite und einen Beispielsatz in die dritte Spalte der Tabelle. Für den Beispielsatz kannst du Sätze aus dem Korpus verwenden, zu denen du gelangst, indem du auf die blau geschriebenen Treffer klickst. Nutze das Online-Wörterbuch PONS (<https://de.pons.com>), falls dir Wörter nicht bekannt sind.

Sustantivo	Adjetivos adecuados	Ejemplo
ventaja	gran(de), mayor, cierta, inmensa, principal, notable, enorme, clara, importante, única, doble, amplia, adicional, aparente, considerable	Estas máquinas presentan <u>ventajas considerables</u> sobre la fabricación manual.
problema	grave, gran(de), serio, principal, mayor, fundamental, importante, numerosos, verdadero, actual, básico, mundial, gravísimo	El <u>problema fundamental</u> que ellos me planteaban era: ¿Cuándo íbamos a hacer la reunión?
objetivo	principal, fundamental, final, primordial, único, común, general, específico, básico, claro, gran, central, verdadero	Su <u>principal objetivo</u> era ver personalmente los daños causados.
hecho	mero, simple, importante, curioso, real	Un par de zapatos viejos, por el <u>simple hecho</u> de haber pertenecido a una estrella de cine, cobran una importancia inusitada.

8. Anhang

precio	alto, gran, buen, menor, mejor, mayor, bajo, exorbitante, ínfimo, inestimable, inferior, excesivo, elevado, irrisorio	En las sociedades occidentales predominan los juguetes electrónicos de <u>precio elevado</u> .
argumento	principal, gran, concluyente, mejor, poderoso, sólido, central, contundente, sencillo, válido, convincente, irrefutable	Otro <u>argumento contundente</u> es el alto nivel de rendimiento.
asunto	grave, principal, importante, delicado, pendiente, urgente, particular, serio, desagradable	¿Dices que tienes que hablarme de un <u>asunto importante</u> ?
razón	sencilla, buena, principal, poderosa, pura, obvia, suprema, clara	¿Por qué se hace esto? – Por la <u>sencilla razón</u> de que no hay textos en Venezuela que se refieran a este tema.
cuestión	grande, grave, importante, pendiente, fundamental, simple, concreta, palpitante, principal	Pero este plan no sirve para resolver la <u>cuestión fundamental</u> .
error	grave, gran(de), gravísimo, funesto, craso, fatal, lamentable, enorme, deplorable, imperdonable	Este <u>lamentable error</u> tenía que suceder más tarde o más temprano.
amigo	buen(o), mejor, íntimo, antiguo, gran(de), fiel, leal, verdadero, digno, entrañable	Giorgio era un <u>verdadero amigo</u> .
idea	nueva, fija, buena, gran(de), vaga, clara, original, central	Es una <u>buena idea</u> todo lo que me ha dicho.
plan	nuevo, ambicioso, estratégico, vasto, concreto, diabólico, divino, futuro, magnífico, detallado	Tengo yo un <u>plan magnífico</u> .

14. Arbeitsblatt 5: *Dichos españoles* – Korpus: CREA (Version 2015)**Dichos españoles****a) Forma los dichos españoles correctos y encuentra el significado adecuado.**

Nutze das *Corpus de Referencia del Español Actual* in der Version von 2015 (<http://corpus.rae.es/creanet.html>) für diese Aufgabe. Zunächst gibst du den zweiten Teil der Redewendung (z. B.: *los platos rotos*) in die Zeile *Consulta* ein und klickst auf *Buscar*. Auf der sich öffnenden Seite klickst du im Bereich *Obtención de ejemplos* auf *Recuperar ,Concordancias‘*. Aus der nun erscheinenden Liste liest du dir einige Sätze durch und wählst den passenden ersten Teil der Redewendung aus (z. B.: *pagar*). Ebenso gut kannst du zwei der untenstehenden Teile, die deiner Meinung nach zusammengehören, gemeinsam eingeben und anhand der Treffer überprüfen, ob diese Kombination korrekt ist. Schließlich wählst du anhand der Beispiele aus dem Korpus die deutsche Bedeutung aus.

CONCORDANCIA

01. Deportivo-Zaragoza (4 - 1). El Zaragoza pagó **los platos rotos** de la eliminación copera del Deportivo de centrocampistas" y negó que ambos pagaran **los platos rotos** de la derrota ante el Atlético de Madrid estrictamente técnicas. Ambos pueden haber pagado **los platos rotos** de la debacle frente al Atlético, a la "solidaridad" al saber la derrota. Nogueira ha pagado **los platos rotos** del cavaquismo al no lograr imponer su enacostumbrada que lo habitual es pagar siempre **los platos rotos** en los conflictos en los que nuestro equipo no a puerto- o serán ellos quien terminen pagando **los platos rotos**. El embajador español en Canadá, José María Laciación de Cataluña La empresa teme acabar pagando **los platos rotos** del conflicto entre la Administración y los sindicatos.

CREA (Version 2015): Konkordanzen zur Suchanfrage *los platos rotos*.

1) estar como	a) por culo	I. in seinem Element sein
2) ser un cero	b) por los codos	II. ein Kinderspiel sein
3) estar en	c) su salsa	III. reden wie ein Wasserfall
4) pagar	d) manitas	IV. seinen Rausch ausschlafen
5) ser	e) la mona	V. umwerfend sein
6) hablar	f) los platos rotos	VI. Du kannst mich mal!
7) vete a tomar	g) la casa por la ventana	VII. Geld zum Fenster hinauswerfen
8) dormir	h) a la izquierda	VIII. eine Niete sein
9) tirar	i) un tren	XI. Händchen halten
10) hacer	j) pan comido	X. die Suppe auslöffeln müssen

b) Escribe cinco frases con los dichos anteriores.

Puedes usar los ejemplos del corpus para esta tarea. Si no conoces algunas palabras, usa el diccionario en línea *PONS* (<https://de.pons.com>).

Dichos españoles – *Solución*

a) Forma los dichos españoles correctos y encuentra el significado adecuado.

Nutze das *Corpus de Referencia del Español Actual* in der Version von 2015 (<http://corpus.rae.es/creanet.html>) für diese Aufgabe. Zunächst gibst du den zweiten Teil der Redewendung (z. B.: *los platos rotos*) in die Zeile *Consulta* ein und klickst auf *Buscar*. Auf der sich öffnenden Seite klickst du im Bereich *Obtención de ejemplos* auf *Recuperar ,Concordancias‘*. Aus der nun erscheinenden Liste liest du dir einige Sätze durch und wählst den passenden ersten Teil der Redewendung aus (z. B.: *pagar*). Ebenso gut kannst du zwei der untenstehenden Teile, die deiner Meinung nach zusammengehören, gemeinsam eingeben und anhand der Treffer überprüfen, ob diese Kombination korrekt ist. Schließlich wählst du anhand der Beispiele aus dem Korpus die deutsche Bedeutung aus.

CONCORDANCIA

01. Deportivo-Zaragoza (4 - 1). El Zaragoza pagó los platos rotos de la eliminación copera del Deportivo de centrocampistas" y negó que ambos pagaran los platos rotos de la derrota ante el Atlético de Madrid estrictamente técnicas. Ambos pueden haber pagado los platos rotos de la debacle frente al Atlético, a la derrota al saber la derrota. Nogueira ha pagado los platos rotos del cavaquismo al no lograr imponer su enacostumbradosa que lo habitual es pagar siempre los platos rotos en los conflictos en los que nuestro an a puerto- o serán ellos quien terminen pagando los platos rotos. El embajador español en Canadá, José Lladro de Cataluña La empresa teme acabar pagando los platos rotos del conflicto entre la Administración y la CREA (Version 2015): Konkordanz zur Suchanfrage *los platos rotos*.

1) estar como	a) por culo	I. in seinem Element sein
2) ser un cero	b) por los codos	II. ein Kinderspiel sein
3) estar en	c) su salsa	III. reden wie ein Wasserfall
4) pagar	d) manitas	IV. seinen Rausch ausschlafen
5) ser	e) la mona	V. umwerfend sein
6) hablar	f) los platos rotos	VI. Du kannst mich mal!
7) vete a tomar	g) la casa por la ventana	VII. Geld zum Fenster hinauswerfen
8) dormir	h) a la izquierda	VIII. eine Niete sein
9) tirar	i) un tren	XI. Händchen halten
10) hacer	j) pan comido	X. die Suppe auslöffeln müssen

Solución: 1i) – V; 2h) – VIII; 3c) – I; 4f) – X; 5j) – II; 6b) – III; 7a) – VI; 8e) – IV; 9g) – VII; 10d) – XI;

b) Escribe cinco frases con los dichos anteriores.

Puedes usar los ejemplos del corpus para esta tarea. Si no conoces algunas palabras, usa el diccionario en línea *PONS* (<https://de.pons.com>).

Lösungsmöglichkeit:

1. Elvira está como un tren. Su novio la quiere mucho y no puede vivir sin ella.
2. Mi cuñado Antonio es un intelectual muy culto y leído, pero para las cuestiones prácticas es un cero a la izquierda.

3. El encuentro, como el anterior, se hace prometedor, porque ambos equipos están en su salsa con el juego rápido y los lanzamientos a los pocos minutos de juego.
4. Si aquí nadie resume las responsabilidades, os hacen pagar los platos rotos.
5. Para un Macintosh el manejar ventanas, iconos, ratón, música y dibujitos en la pantalla es pan comido.
6. Javier estuvo realmente simpático, hablando hasta por los codos y haciendo toda clase de bromas.
7. Anda, tío, pues vete. Vete a tomar por culo de aquí, que no te quiero ni ver.
8. Tu hermana está durmiendo la mona dos cuartos más allá. La estoy oyendo roncar.
9. No estamos como para tirar la casa por la ventana gastándonos el dinero que no tenemos en comer fuera y en irnos de tapeo.
10. Todos admiraban a Barbara Bach, la actual esposa de Ringo, que hacía manitas con su marido.

15. Arbeitsblatt 6: *Sinónimos para adjetivos corrientes* – CdE:Orig (Version 2008)

Sinónimos para adjetivos corrientes

Verwende die spanische Version des *Corpus del Español: Género/Histórico* mit dem alten Interface von 2008, um synonyme Adjektive zu ermitteln (<https://www.corpusdelespanol.org/hist-gen/2008/x.asp>). Zunächst gibst du das jeweilige Substantiv in eckigen Klammern in die Zeile *Palabra(s)* ein, klickst auf das Fragezeichen in der Zeile *Colocados* und trägst das Adjektiv, für welches ein Synonym gesucht werden soll, in eckigen Klammern mit einem vorausgehenden Gleichheitszeichen dort ein. Daraufhin bestimmst du die Kollokationsspanne mit jeweils einem Wort links und rechts des Suchbegriffs. Schließlich wählst du im Bereich *Ordenar* die Option *Relevancia* aus und im Bereich *Mínimo* die Option *Frecuencia*, 3'. Auf diese Weise werden die Adjektive nach deren Relevanz für das Substantiv ausgewählt und die jeweilige Kombination muss mindestens dreimal im Korpus erscheinen. Nach einem Klick auf das Feld *Buscar* wird eine Liste mit Treffern angezeigt, aus der du einige Adjektive in die zweite und einen Beispielsatz in die dritte Spalte der Tabelle schreibst. Für den Beispielsatz kannst du Sätze aus dem Korpus verwenden, zu denen du gelangst, indem du auf die blau geschriebenen Treffer klickst. Nutze das Online-Wörterbuch PONS (<https://de.pons.com>), falls dir Wörter nicht bekannt sind.

Interface des CdE:Orig (Version 2008).

Adjetivo “corriente”	Adjetivos casi sinónimos	Ejemplo
ventaja “grande”	enorme, importante, mayor	Barcelona tiene una <u>enorme ventaja</u> : el mar.
problema “grande”		
objetivo “importante”		

hecho “importante”		
precio “alto”		
argumento “importante”		
asunto “importante”		
plan “bueno”		
cuestión “importante”		

Sinónimos para adjetivos corrientes – *Solución*

Verwende die spanische Version des *Corpus del Español: Género/Histórico* mit dem alten Interface von 2008, um synonyme Adjektive zu ermitteln (<https://www.corpusdelespanol.org/hist-gen/2008/x.asp>). Zunächst gibst du das jeweilige Substantiv in eckigen Klammern in die Zeile *Palabra(s)* ein, klickst auf das Fragezeichen in der Zeile *Colocados* und trägst das Adjektiv, für welches ein Synonym gesucht werden soll, in eckigen Klammern mit einem vorausgehenden Gleichheitszeichen dort ein. Daraufhin bestimmst du die Kollokationsspanne mit jeweils einem Wort links und rechts des Suchbegriffs. Schließlich wählst du im Bereich

Interface des CdE:Orig (Version 2008).

Ordenar die Option *Relevancia* aus und im Bereich *Mínimo* die Option *Frecuencia* ,3'. Auf diese Weise werden die Adjektive nach deren Relevanz für das Substantiv ausgewählt und die jeweilige Kombination muss mindestens dreimal im Korpus erscheinen. Nach einem Klick auf das Feld *Buscar* wird eine Liste mit Treffern angezeigt, aus der du einige Adjektive in die zweite und einen Beispielsatz in die dritte Spalte der Tabelle schreibst. Für den Beispielsatz kannst du Sätze aus dem Korpus verwenden, zu denen du gelangst, indem du auf die blau geschriebenen Treffer klickst. Nutze das Online-Wörterbuch PONS (<https://de.pons.com>), falls dir Wörter nicht bekannt sind.

Adjetivo “corriente”	Adjetivos casi sinónimos	Ejemplo
ventaja “grande”	enorme, importante, mayor	Barcelona tiene una <u>enorme ventaja</u> : el mar.
problema “grande”	tremendo, importante, mayor, enorme	El <u>mayor problema</u> es la incomunicación.
objetivo “importante”	primordial, estratégico, principal, fundamental, esencial	El <u>principal objetivo</u> de este proyecto es la construcción de una escuela para huérfanos.
hecho “importante”	significativo, fundamental, notable, grave, grande	Se ignoraba un <u>hecho significativo</u> que había ocurrido aquella misma noche.

precio “alto”	elevado, caro, superior	No nos importa pagar el <u>elevado precio</u> .
argumento “importante”	sólido, poderoso, principal	Ella supo que no tenía ningún <u>argumento sólido</u> para contestarle.
asunto “importante”	primordial, urgente, transcendental, grave, vital, principal, serio, fundamental, interesante	Por fin se decidía a abordar un <u>asunto serio</u> .
plan “bueno”	magnífico, propio	Tengo yo un <u>plan magnífico</u> .
cuestión “importante”	grave, fundamental, interesante, principal	La <u>cuestión principal</u> no es tanto el peligro actual como el peligro para las generaciones futuras.

16. Arbeitsblatt 7: *Homónimos* – Korpus: CdE:Nuevo**Homónimos**

Nutze das *Corpus del Español: Web/Dialectos* (CdE:Nuevo), um das korrekte Genus für die folgenden Homonyme zu ermitteln (<https://www.corpusdelespanol.org/web-dial/>). Zunächst wählst du die Funktion *Colocados* aus, gibst den Suchbegriff (z. B.: *cabeza*) in die Zeile *Palabra/frase* ein, bestimmst *art.Todos* als Wortart im Bereich *Colocados* sowie ein Wort links des Suchbegriffs als Intervall. Nach einem Klick auf das Feld *Buscar* wird eine Liste mit Artikeln angezeigt, aus der du einen maskulinen oder femininen Artikel auswählst. Daraufhin klickst du auf *Contexto*, liest dir einige Beispielsätze durch und entscheidest, ob es sich dabei um die erste oder zweite der unten angegebenen Bedeutungen handelt. Markiere das Feld mit der jeweiligen deutschen Bedeutung in roter Farbe (für feminine Wörter) oder blauer Farbe (für maskuline Wörter). Schließlich wählst du diejenige Farbe für das Feld mit dem Pluszeichen aus, deren Bedeutung häufiger im Korpus erscheint.

[Lista](#)
[Gráfico](#)
[Colocados](#)
[Comparar](#)
[PCEC](#)

Palabra/frase [POS]

Colocados

CONTEXTO	FREC
LA	175246
UNA	4122
EL	887
UN	371
LOS	94
LAS	60
LO	19
UNOS	10
TOTAL	180809

Links: Suche nach dem Substantiv *cabeza* mit einem Artikel; Rechts: Artikel, die mit dem Substantiv *cabeza* erscheinen.

Palabra española	Significado 1	más usado	Significado 2
el / la cabeza	der Kopf		der Anführer, Leiter
el/ la capital	das Kapital		die Hauptstadt
el / la policía	die Polizei (die Polizistin)		der Polizist
el / la cura	der Pfarrer		die Kur
el / la guía	der (An-) Führer		die Führung, das Handbuch [die (An-) Führerin]
el / la rosa	die Rose		die Farbe Rosa
el / la coma	das Komma		das Koma
el / la cometa	der Komet		der Drachen (Spielzeug)
el / la cólera	der Zorn		die Cholera
el / la papa	die Kartoffel		der Papst
el / la orden	die Ordnung		der Befehl

Homónimos – Solución

Nutze das *Corpus del Español: Web/Dialectos* (CdE:Nuevo), um das korrekte Genus für die folgenden Homonyme zu ermitteln (<https://www.corpusdelespanol.org/web-dial/>). Zunächst wählst du die Funktion *Colocados* aus, gibst den Suchbegriff (z. B.: *cabeza*) in die Zeile *Palabra/frase* ein, bestimmst *art.Todos* als Wortart im Bereich *Colocados* sowie ein Wort links des Suchbegriffs als Intervall. Nach einem Klick auf das Feld *Buscar* wird eine Liste mit Artikeln angezeigt, aus der du einen maskulinen oder femininen Artikel auswählst. Daraufhin klickst du auf *Contexto*, liest dir einige Beispielsätze durch und entscheidest, ob es sich dabei um die erste oder zweite der unten angegebenen Bedeutungen handelt. Markiere das Feld mit der jeweiligen deutschen Bedeutung in roter Farbe (für feminine Wörter) oder blauer Farbe (für maskuline Wörter). Schließlich wählst du diejenige Farbe für das Feld mit dem Pluszeichen aus, deren Bedeutung häufiger im Korpus erscheint.

[Lista](#)
[Gráfico](#)
[Colocados](#)
[Comparar PCEC](#)

Palabra/frase [POS]

Colocados

CONTEXTO	FREC
LA	175246
UNA	4122
EL	887
UN	371
LOS	94
LAS	60
LO	19
UNOS	10
TOTAL	180809

Links: Suche nach dem Substantiv *cabeza* mit einem Artikel; Rechts: Artikel, die mit dem Substantiv *cabeza* erscheinen.

Palabra española	Significado 1	más usado	Significado 2
el / la cabeza	der Kopf		der Anführer, Leiter
el/ la capital	das Kapital		die Hauptstadt
el / la policía	die Polizei (die Polizistin)		der Polizist
el / la cura	der Pfarrer		die Kur
el / la guía	der (An-) Führer		die Führung, das Handbuch [die (An-) Führerin]
el / la rosa	die Rose		die Farbe Rosa
el / la coma	das Komma		das Koma
el / la cometa	der Komet		der Drachen (Spielzeug)
el / la cólera	der Zorn		die Cholera
el / la papa	die Kartoffel		der Papst
el / la orden	die Ordnung		der Befehl

17. Arbeitsblatt 8: *Formación de palabras* – Korpus: CREA**Formación de palabras****1. Formación de palabras con prefijos**

Combina los adjetivos siguientes con los prefijos adecuados para formar lo contrario.

Nutze die annotierte Version des *Corpus de Referencia del Español Actual* (CREA), um die korrekten Wörter zu finden (<http://web.frl.es/CREA/view/inicioExterno.view>). Zunächst gibst du die Platzhalter '?*' zusammen mit dem Adjektiv (z. B.: *justo*) als *Lema* in der Benutzeroberfläche des Korpus ein und wählst die jeweilige Wortart aus (z. B.: *adjetivo*). Nach einem Klick auf das Feld *Concordancia* wird eine Liste mit Sätzen angezeigt, die ein entsprechendes Adjektiv enthalten. Daraufhin schreibst du das Adjektiv in die passende Spalte.¹³³³

Corpus de Referencia del Español Actual
(CREA. Versión anotada)

REAL ACADEMIA ESPAÑOLA

Concordancias | Coapariciones | Configuración | Ayuda | Modo de cita | Sugerencias

Lema: ?*justo Forma: Clase de palabra: adjetivo ☐ Grafía original + Subcorpus

Número: Género: Tipo: Grado:

2.506 casos en 1.681 documentos.

REF. (Clasificación, país)	CONCORDANCIA	Ordenar por:
1 1975 Esp.	Las protestas contra la ley injusta	
2 1975 Cuba	el arte de los africanos, se ha mantenido un "desprecio ridículo y profundamente injusto ". Se sabe desde hace si	
3 1975 Arg.	éxito, tendrán que reservar los platos suculentos para comerlos afuera: es penoso e injusto y contraproducente proh	
4 1975 Esp.	— Todo ha terminado mal porque la vida es muy injusta con las personas de bue	
5 1975 Esp.	No es falta de santidad, sino sobra de dolores injustos .	
6 1975 Perú	4. Finalmente queremos expresar que a pesar de los ataques injustos y mal intencionados que	
7 1975 Esp.	en su calidad de capitán del equipo, dialogaba con el árbitro, fue una decisión injusta . A los tres días el presid	

CREA: Suche nach dem richtigen Präfix zur Bildung des Gegenteils von *justo*.

→ justo	suficiente	posible	preciso	conectado
social	variable	ordenado	agradable	comunista
adecuado	racista	perfecto	legal	paciente
				conocido

anti-	des-	in-	im-	i-
		injusto		

→ Prefijos no cambian la parte de la oración.

¹³³³ Der Platzhalter '?*' bedeutet *genau ein Zeichen* und der Platzhalter '?*' bedeutet *kein, ein oder mehr als ein Zeichen*. Auf diese Weise werden Wörter gesucht, die aus dem ursprünglichen Adjektiv und einem Präfix bestehen, ohne jedoch die genaue Anzahl der Buchstaben des Präfixes angeben zu müssen.

2. Formación de palabras con sufijos

Combina las palabras a la izquierda con el sufijo adecuado para formar adjetivos.

Verwende dafür das *Corpus de Referencia del Español Actual* (CREA), wie in der vorherigen Aufgabe. Zunächst gibst du den Wortstamm mit den beiden Platzhaltern '?*' (z. B.: *energ?**) als *Lema* ein und wählst dann die gesuchte Wortart aus (z. B.: *adjetivo*). Nach einem Klick auf das Feld *Concordancia* erscheint eine Liste mit Beispielen, denen du das passende Adjektiv entnehmen kannst.

sustantivo o verbo	+ sufijo	= adjetivo
→ energía religión aceptar dolor vulnerar	-able	
economía recomendar islam fantasía comparar	-ico	enérgico,
orgullo lamentar nervio peligro democracia	-oso	

Haz lo mismo con las palabras siguientes para formar sustantivos. ¡Ojo! A veces hay más de una solución.

sustantivo, adjetivo o verbo	+ sufijo	= sustantivo
→ limpio inmigrar seguro	-ción	
pobre visitar explicar	-ez / -eza	limpieza,
crecer sentir real	-ante	
triste vestir probable preparar	-idad	
participar conocer	-tor / -dor	
producir pescar	-miento	
activo rápido mover	-ado / -ido	


→ Sufijos cambian la parte de la oración.

Formación de palabras – Solución

1. Formación de palabras con prefijos

Combina los adjetivos siguientes con los prefijos adecuados para formar lo contrario.

Nutze die annotierte Version des *Corpus de Referencia del Español Actual* (CREA), um die korrekten Wörter zu finden (<http://web.frl.es/CREA/view/inicioExterno.view>). Zunächst gibst du die Platzhalter '?*' zusammen mit dem Adjektiv (z. B.: *justo*) als *Lema* in der Benutzeroberfläche des Korpus ein und wählst die jeweilige Wortart aus (z. B.: *adjetivo*). Nach einem Klick auf das Feld *Concordancia* wird eine Liste mit Sätzen angezeigt, die ein entsprechendes Adjektiv enthalten. Daraufhin schreibst du das Adjektiv in die passende Spalte.¹³³⁴

 **Corpus de Referencia del Español Actual**
(CREA. Versión anotada)

REAL ACADEMIA ESPAÑOLA

Concordancias | **Coapariciones** | **Configuración** | **Ayuda** | **Modo de cita** | **Sugerencias**

Lema: ?*justo Forma: Clase de palabra: adjetivo ☐ Grafía original + Subcorpus

Número: Género: Tipo: Grado:

2.506 casos en 1.681 documentos.

REF. (Clasificación, país)	CONCORDANCIA	Ordenar por:
1 1975 Esp.	Las protestas contra la ley injusta	
2 1975 Cuba	el arte de los africanos, se ha mantenido un "desprecio ridículo y profundamente injusto ". Se sabe desde hace si	
3 1975 Arg.	éxito, tendrán que reservar los platos suculentos para comerlos afuera: es penoso e injusto y contraproducente proh	
4 1975 Esp.	— Todo ha terminado mal porque la vida es muy injusta con las personas de bu	
5 1975 Esp.	No es falta de santidad, sino sobra de dolores injustos .	
6 1975 Perú	4. Finalmente queremos expresar que a pesar de los ataques injustos y mal intencionados que	
7 1975 Esp.	en su calidad de capitán del equipo, dialogaba con el árbitro, fue una decisión injusta . A los tres días el presi	

CREA: Suche nach dem richtigen Präfix zur Bildung des Gegenteils von *justo*.

➔ justo	suficiente	posible	preciso	conectado
	social	variable	ordenado	agradable
				comunista
adecuado	racista	perfecto	legal	paciente
				conocido

anti-	des-	in-	im-	i-
antisocial	desconectado	injusto	imposible	ilegal
anticomunista	desordenado	insuficiente	impreciso	
antirracista	desagradable	invariable	imperfecto	
	desconocido	inadecuado	impaciente	

➔ **Prefijos no cambian la parte de la oración.**

¹³³⁴ Der Platzhalter „?“ bedeutet *genau ein Zeichen* und der Platzhalter „*“ bedeutet *kein, ein oder mehr als ein Zeichen*. Auf diese Weise werden Wörter gesucht, die aus dem ursprünglichen Adjektiv und einem Präfix bestehen, ohne jedoch die genaue Anzahl der Buchstaben des Präfixes angeben zu müssen.

2. Formación de palabras con sufijos

Combina las palabras a la izquierda con el sufijo adecuado para formar adjetivos.

Verwende dafür das *Corpus de Referencia del Español Actual* (CREA), wie in der vorherigen Aufgabe. Zunächst gibst du den Wortstamm mit den beiden Platzhaltern '?*' (z. B.: *energ?**) als *Lema* ein und wählst dann die gesuchte Wortart aus (z. B.: *adjetivo*). Nach einem Klick auf das Feld *Concordancia* erscheint eine Liste mit Beispielen, denen du das passende Adjektiv entnehmen kannst.

sustantivo o verbo	+ sufijo	= adjetivo
→ energía religión aceptar dolor vulnerar	-able	aceptable, vulnerable, recomendable, comparable, lamentable
economía recomendar islam fantasía comparar	-ico	enérgico, económico, islámico, fantástico, democrático
orgullo lamentar nervio peligro democracia	-oso	religioso, doloroso, orgulloso, nervioso, peligroso

Haz lo mismo con las palabras siguientes para formar sustantivos. ¡Ojo! A veces hay más de una solución.

sustantivo, adjetivo o verbo	+ sufijo	= sustantivo
→ limpio inmigrar seguro	-ción	inmigración, explicación, preparación, participación, producción
pobre visitar explicar	-ez / -eza	limpieza, pobreza, tristeza, rapidez
crecer sentir real	-ante	inmigrante, visitante, participante
triste vestir probable preparar	-idad	seguridad, realidad, probabilidad, actividad
participar conocer	-tor / -dor	productor, pescador
producir pescar	-miento	crecimiento, conocimiento, movimiento
activo rápido mover	-ado / -ido	sentido, vestido, conocido, pescado

→ Sufijos cambian la parte de la oración.

18. Arbeitsblatt 9: *Familias de palabras* – Korpus: CORPES**Familias de palabras**

Rellena la tabla con el resto de las palabras pertenecientes a la familia.

Verwende dafür das *Corpus del Español del Siglo XXI* – CORPES (<http://web.frl.es/CORPES/view/inicioExterno.view>). Als *Lema* gibst du den Wortstamm zwischen zwei Asterisken ein (z. B.: **segur**), und im Bereich *Clase de palabra* wählst du die Wortart des gesuchten Worts aus (z. B.: *verbo*).¹³³⁵ Zusätzlich kannst du die Angabe *infinitivo* im Bereich *Tiempo* auswählen. Nach einem Klick auf das Feld *Concordancia* erscheinen einige Beispiele mit dem Verb *asegurar*, welches du als Lösung in die Tabelle eintragen kannst.

Concordancias | Coapariciones | Configuración | Ayuda | Modo de cita | Sugerencias

Lema: **segur** Forma: _____ Clase de palabra: verbo ☐ Grafía original + Subcorpus + Proximidad

Modo: Tiempo: infinitivo Persona: Número: Tipo:

+ -

14.130 casos en 9.042 documentos.

REF. (Clasificación, país)	CONCORDANCIA	Ordenar por: Año ascendente
1 2001 Chile	Primero me llevaron a los sótanos de un hospital aunque a ciencia cierta no podría asegurar que aquello fuera un hospital, en donde un	
2 2001 Chile	grande es el daño, al parecer es imposible rastrear sus pasos en la red. Nadie puede asegurar que la situación está bajo control. Cualquiera	
3 2001 Ven.	obtenía facilidades de fijar los precios en las mercancías de la provincia y de asegurar sus ganancias en España.	
4 2001 Ven.	participantes entre la tierra firme y las embarcaciones respectivamente, con el fin de asegurar el éxito de la insurrección y garantizar la de	
5 2001 Col.	las artes de las conquistas y el amor. Así que te puedes quedar tranquila. Puedo asegurar te con certeza que nunca gozó de sus favor	

CORPES: Suche nach einem mit dem Adjektiv *seguro* verwandten Verb.

sustantivo	verbo	adjetivo
	<i>asegurar</i>	seguro/a
la igualdad		
		peor
		gordo/a
la familia		
	demostrar	
		mejor
	herir	
		cansado/a
	tolerar	
		posible
la motivación		
	depender	
		formal
	olvidar	
		normal
		regular
	aparecer	

¹³³⁵ Der Platzhalter **,** bedeutet *kein, ein oder mehr als ein Zeichen*. Auf diese Weise werden Wörter gesucht, die aus dem Wortstamm und einem optionalen Prä- und/oder Suffix variabler Länge bestehen.

Familias de palabras – Solución

Rellena la tabla con el resto de las palabras pertenecientes a la familia.

Verwende dafür das *Corpus del Español del Siglo XXI* – CORPES (<http://web.frl.es/CORPES/view/inicioExterno.view>). Als *Lema* gibst du den Wortstamm zwischen zwei Asterisken ein (z. B.: **segur**), und im Bereich *Clase de palabra* wählst du die Wortart des gesuchten Worts aus (z. B.: *verbo*).¹³³⁶ Zusätzlich kannst du die Angabe *infinitivo* im Bereich *Tiempo* auswählen. Nach einem Klick auf das Feld *Concordancia* erscheinen einige Beispiele mit dem Verb *asegurar*, welches du als Lösung in die Tabelle eintragen kannst.

Concordancias | Coapariciones | Configuración | Ayuda | Modo de cita | Sugerencias

Lema: Forma: Clase de palabra: ☐ Grafía original

Modo: Tiempo: Persona: Número: Tipo:

14.130 casos en 9.042 documentos.

REF. (Clasificación, país)	CONCORDANCIA	Ordenar por: Año ascendente
1 2001 Chile	Primero me llevaron a los sótanos de un hospital aunque a ciencia cierta no podría asegurar que aquello fuera un hospital, en donde un	
2 2001 Chile	grande es el daño, al parecer es imposible rastrear sus pasos en la red. Nadie puede asegurar que la situación está bajo control. Cualquiera	
3 2001 Ven.	obtenía facilidades de fijar los precios en las mercancías de la provincia y de asegurar sus ganancias en España.	
4 2001 Ven.	participantes entre la tierra firme y las embarcaciones respectivamente, con el fin de asegurar el éxito de la insurrección y garantizar la de	
5 2001 Col.	las artes de las conquistas y el amor. Así que te puedes quedar tranquila. Puedo asegurar te con certeza que nunca gozó de sus favor	

CORPES: Suche nach einem mit dem Adjektiv *seguro* verwandten Verb.

sustantivo	verbo	adjetivo
la seguridad	<i>asegurar</i>	seguro/a
la igualdad	igualar	igual
el empeoramiento	empeorar	peor
la gorda, el gordo	engordar	gordo/a
la familia	familiarizar	familiar
la demostración	demostrar	demostrable, demostrativo/a
la mejora, la mejoría, el mejoramiento	mejorar	mejor
la herida	herir	herido/a
el cansancio	cansar	cansado/a
la tolerancia	tolerar	tolerante, tolerable
la posibilidad	posibilitar	posible
la motivación	motivar	motivante, motivador
la (in-)dependencia, el/la dependiente/a	depender	(in-)dependiente
el (in-)formalismo	formalizar	formal
el olvido	olvidar	(in-)olvidable
la normalidad	normalizar	normal
la (ir)regularidad	regular, regularizar	regular
la (des-)aparición	aparecer	aparente

¹³³⁶ Der Platzhalter *,* bedeutet *kein, ein oder mehr als ein Zeichen*. Auf diese Weise werden Wörter gesucht, die aus dem Wortstamm und einem optionalen Prä- und/oder Suffix variabler Länge bestehen.

19. Arbeitsblatt 10: *Familias de palabras: La dictadura* – Korpus: CORPES**Familias de palabras: *La dictadura***

Rellena la tabla con el resto de las palabras pertenecientes a la familia.

Verwende dafür das *Corpus del Español del Siglo XXI* – CORPES (<http://web.frl.es/CORPES/view/inicioExterno.view>). Als *Lema* gibst du den Wortstamm zwischen zwei Asterisken ein (z. B.: **cautiv**), und im Bereich *Clase de palabra* wählst du die Wortart des gesuchten Worts aus (z. B.: *verbo*).¹³³⁷ Zusätzlich kannst du die Angabe *infinitivo* im Bereich *Tiempo* auswählen. Nach einem Klick auf das Feld *Concordancia* erscheinen einige Beispiele mit dem Verb *cautivar*, welches du als Lösung in die Tabelle eintragen kannst.

Concordancias | Coapariciones | Configuración | Ayuda | Modo de cita | Sugerencias

Lema: **cautiv** Forma: Clase de palabra: verbo Grafiá original: Subcorpus: Proximidad: Modo: Tiempo: infinitivo Persona: Número: Tipo:

Concordancia Estadística Nueva consulta

349 casos en 322 documentos.

REF. (Clasificación, país) CONCORDANCIA Ordenar por: Año ascendente

1	2001 Arg.	convencida de que los celos y la competencia con un amante eran el mejor recurso para cautivar a un hombre. Al fallarle, debía sentir que también en eso
2	2001 Arg.	proyectos artísticos parecían integrar una misma parte de sus vidas. Exhibirse, cautivar al otro en lo que creen es su originalidad, su distinción, in
3	2001 Arg.	incontenible verborragia sobre los mitos y creencias taoístas; curiosamente lograba cautivar a muchos.
4	2001 Hond.	De vez en cuando -a fuerza de su persistencia- logra alguna frase capaz de cautivar a muchos, circunstancia que no deja ir en vano y aprovecha
5	2001 Col.	actitud despectiva sobre todas este tipo de expresiones culturales, que antes que cautivar los los atemorizaban.

CORPES: Suche nach einem mit dem Substantiv *cautiverio* verwandten Verb.

sustantivo	verbo	adjetivo
el cautiverio	<i>cautivar</i>	
		(i)legítimo/a
	asesinar	
la dictadura		
	sospechar	
el golpe (de Estado)		
	sufrir	
el exilio		
		subversivo/-a
el juicio		
		militar
	combatir	
la tortura		
	violar	
la venganza		
		desaparecido/-a
la guerra		
		(in-)justo/-a

¹³³⁷ Der Platzhalter *,* bedeutet *kein, ein oder mehr als ein Zeichen*. Auf diese Weise werden Wörter gesucht, die aus dem Wortstamm und einem optionalen Prä- und/oder Suffix variabler Länge bestehen.

Familias de palabras: *La dictadura – Solución*

Rellena la tabla con el resto de las palabras pertenecientes a la familia.

Verwende dafür das *Corpus del Español del Siglo XXI* – CORPES (<http://web.frl.es/CORPES/view/inicioExterno.view>). Als *Lema* gibst du den Wortstamm zwischen zwei Asterisken ein (z. B.: **cautiv**), und im Bereich *Clase de palabra* wählst du die Wortart des gesuchten Worts aus (z. B.: *verbo*).¹³³⁸ Zusätzlich kannst du die Angabe *infinitivo* im Bereich *Tiempo* auswählen. Nach einem Klick auf das Feld *Concordancia* erscheinen einige Beispiele mit dem Verb *cautivar*, welches du als Lösung in die Tabelle eintragen kannst.

Concordancias | Coapariciones | Configuración | Ayuda | Modo de cita | Sugerencias

Lema: **cautiv** Forma: Clase de palabra: verbo ☐ Grafía original + Subcorpus + Proximidad

Modo: Tiempo: infinitivo Persona: Número: Tipo:

+ - Concordancia Estadística Nueva consulta

349 casos en 322 documentos.

REF. (Clasificación, país) CONCORDANCIA Ordenar por: Año ascendente

1	2001 Arg.	convencida de que los celos y la competencia con un amante eran el mejor recurso para	cautivar	a un hombre. Al fallarle, debía sentir que también en eso
2	2001 Arg.	proyectos artísticos parecían integrar una misma parte de sus vidas. Exhibirse,	cautivar	al otro en lo que creen es su originalidad, su distinción, in
3	2001 Arg.	incontenible verborragia sobre los mitos y creencias taoístas; curiosamente lograba	cautivar	a muchos.
4	2001 Hond.	De vez en cuando -a fuerza de su persistencia- logra alguna frase capaz de	cautivar	a muchos, circunstancia que no deja ir en vano y aprovecha
5	2001 Col.	actitud despectiva sobre todas este tipo de expresiones culturales, que antes que	cautivar	los los atemorizaban.

CORPES: Suche nach einem mit dem Substantiv *cautiverio* verwandten Verb.

sustantivo	verbo	adjetivo
el cautiverio	<i>cautivar</i>	cautivo/-a
la (i)legitimidad, la legitimación	legitimar	(i)legítimo/a
el asesinato, el/la asesino/-a	asesinar	asesino/-a
la dictadura	dictar	dictatorial
la sospecha, el/la sospechoso/-a	sospechar	sospechoso/-a
el golpe (de Estado)	golpear	golpista
el sufrimiento	sufrir	(in)sufrible
el exilio	exiliarse	exiliado/-a
la subversión	subvertir	subversivo/-a
el juicio	enjuiciar	juicioso/-a
el/la militante	militarizar	militar
el combate, el/la combatiente	combatir	combativo/-a, combatiente
la tortura	torturar	torturador/-a, torturante
la violación, el/la violador/-a	violar	violento/-a, inviolable
la venganza	vengar	vengador/-a, vengativo/-a
la desaparición	desaparecer	desaparecido/-a
la guerra	guerrear	guerrero/-a, guerrillero/-a
la (in-)justicia, la justificación	justificar	(in-)justo/-a

¹³³⁸ Der Platzhalter *,* bedeutet *kein, ein oder mehr als ein Zeichen*. Auf diese Weise werden Wörter gesucht, die aus dem Wortstamm und einem optionalen Prä- und/oder Suffix variabler Länge bestehen.

20. Arbeitsblatt 11: *El alemán 'werden'* – Korpus: CREA / CREA (Version 2015)**El alemán 'werden'**

- a) ¿Qué equivalente para el alemán *werden* necesitamos en qué contexto? Busca el verbo correcto para cada frase y ponlo en el tiempo y la persona correctos.

Verwende das *Corpus de Referencia del Español Actual* – CREA (Version 2015) für diese Aufgabe (<http://corpus.rae.es/creanet.html>). Im Feld *Consulta* gibst du eines der untenstehenden Verben zusammen mit dem Adjektiv oder Substantiv ein, das ihm im jeweiligen Satz folgt (z. B.: *hacerse famosa*), und klickst auf *Buscar*. Auf der sich öffnenden Seite siehst du, ob es für diese Kombination Treffer gibt oder nicht, und findest auf diese Weise ein passendes Verb für jeden Satz. Durch einen Klick auf *Recuperar 'Concordancias'* im Bereich *Obtención de ejemplos* gelangst du zu den Konkordanzen.

Real Academia Española - Corpus de Referencia del Español Actual (CREA)

Consulta:	<input type="text" value="hacerse famosa"/>		
Criterios de selección:			
Autor:	<input type="text"/>	Obra:	<input type="text"/>

CREA (Version 2015): Suche nach der Kombination *hacerse + famosa*.

hacerse	llegar a ser	volverse
quedarse	convertirse	
ponerse	querer ser	meterse

- Una actriz _____ famosa cuando interpreta un papel importante en una película.
- Juan _____ rojo como un tomate.
- Un diabético puede _____ ciego por tener el azúcar alto.
- Su ex-mujer _____ loca.
- José Coromina _____ profesor y subdirector.
- Tenía que _____ en musulmán a causa del cambio de su vida.
- Desde la infancia quería _____ monja.
- Luis _____ escritor, pero a su madre no le gusta esa idea.
- ¿Porqué canta ese gallo cuando _____ de noche?
- El acusado puede _____ agresivo y tener reacciones paradójicas de "huida hacia adelante".
- Ana _____ embarazada de Miguel cuando tenía 19 años.
- La madre empezó a _____ furiosa a causa de todo el desorden en el cuarto del niño.

b) ¿Qué diferencia hay entre los verbos del ejercicio a)? Rellena la tabla siguiente.¹³³⁹

¿Con qué clase de palabra pueden ser combinados los equivalentes del alemán *werden*?

significado	equivalente de <i>werden</i>	+	clase de palabra
Grundlegende Veränderung / Wandlung (auch Berufe)	_____ en / meterse	+	_____
Langsame Entwicklung / allmähliche Veränderung	_____	+	adjetivo / _____
Plötzliche, überraschende, starke Gemütsveränderung	volverse	+	_____/sustantivo
Endgültige Veränderung / Ergebnis	quedarse	+	_____
Vorübergehender Zustand / Veränderung	_____	+	adjetivo
Absicht / Vorhaben	_____	+	adjetivo/_____
Erreichen eines Ziels	llegar a ser	+	adjetivo/_____

c) También hay palabras propias para expresar el alemán *werden*.

Verwende die neue, annotierte Version des CREA, um die korrekten Verben zu finden, wie im untenstehenden Beispiel (<http://web.frl.es/CREA/view/inicioExterno.view>). Als *Lema* gibst du den Wortstamm zwischen zwei Asterisken ein (z. B.: **peor**) und wählst *verbo* im Bereich *Clase de palabra* sowie *infinitivo* im Bereich *Tiempo* aus.¹³⁴⁰ Nach einem Klick auf das Feld *Concordancia* erscheint eine Liste mit Beispielsätzen, an denen du das korrekte Verb ablesen kannst.

Concordancias | Coapariciones | Configuración | Ayuda | Modo de cita | Sugerencias

Lema: Forma: Clase de palabra: ☐ Grafía original

Modo: Tiempo: Persona: Número: Tipo:

265 casos en 234 documentos.

REF. (Clasificación, país)	CONCORDANCIA	Ordenar por
1 1975 Méx.	Lupita: (Sentenciosa.) Siempre que algo cambia es para empeorar .	
2 1975 Esp.	Tramacho moviendo la cabeza dubitativamente: «Falsas ilusiones, amigo: no tardará en empeorar para siempre. He visto	
3 1975 Arg.	no recurran jamás a los retos, humillaciones, castigos y burlas que sólo han de empeorar la situación.	
4 1976 Ven.	podrían utilizar las drogas como tranquilizantes. Sin embargo las mismas drogas pueden empeorar su condición mental, e	
5 1976 Ven.	interpretada como empeoramiento del cuadro esquizofrénico, aumentando la dosis para empeorar todavía más los síntom	

CREA: Suche nach einem mit dem Adjektiv *peor* verwandten Verb.

schlechter werden: empeorar

krank werden: _____

besser werden: _____

traurig werden: _____

dick werden: _____

Nacht werden: _____

schlank werden: _____

dunkel werden: _____

d) Ahora escribe una frase para cada verbo del ejercicio c) en tu cuaderno de ejercicios. Puedes utilizar el CREA para encontrar un contexto apropiado.

¹³³⁹ Einteilung in Anlehnung an: Görrissen (2005): op. cit., S. 220-221. Und De Bruyne (2002): op. cit., S. 578-583.

¹³⁴⁰ Der Platzhalter *,* bedeutet *kein, ein oder mehr als ein Zeichen*. Auf diese Weise werden Wörter gesucht, die aus dem Wortstamm und einem optionalen Prä- und/oder Suffix variabler Länge bestehen.

El alemán ‘werden’ – Solución

a) ¿Qué equivalente para el alemán *werden* necesitamos en qué contexto? Busca el verbo correcto para cada frase y ponlo en el tiempo y la persona correctos.

Verwende das *Corpus de Referencia del Español Actual* – CREA (Version 2015) für diese Aufgabe (<http://corpus.rae.es/creanet.html>). Im Feld *Consulta* gibst du eines der untenstehenden Verben zusammen mit dem Adjektiv oder Substantiv ein, das ihm im jeweiligen Satz folgt (z. B.: *hacerse famosa*), und klickst auf *Buscar*. Auf der sich öffnenden Seite siehst du, ob es für diese Kombination Treffer gibt oder nicht, und findest auf diese Weise ein passendes Verb für jeden Satz. Durch einen Klick auf *Recuperar ‚Concordancias‘* im Bereich *Obtención de ejemplos* gelangst du zu den Konkordanzen.

Real Academia Española - Corpus de Referencia del Español Actual (CREA)

Consulta:	<input type="text" value="hacerse famosa"/>		
Criterios de selección:			
Autor:	<input type="text"/>	Obra:	<input type="text"/>

CREA (Version 2015): Suche nach der Kombination *hacerse* + *famosa*.

hacerse		llegar a ser		volverse
	quedarse		convertirse	
ponerse		querer ser		meterse

1. Una actriz se hace famosa cuando interpreta un papel importante en una película.
2. Juan se puso rojo como un tomate.
3. Un diabético puede quedarse ciego por tener el azúcar alto.
4. Su ex-mujer se ha vuelto loca.
5. José Coromina llegó a ser profesor y subdirector.
6. Tenía que convertirse en musulmán a causa del cambio de su vida.
7. Desde la infancia quería meterse monja.
8. Luis quiere ser escritor, pero a su madre no le gusta esa idea.
9. ¿Porqué canta ese gallo cuando se hace de noche?
10. El acusado puede volverse agresivo y tener reacciones paradójicas de “huida hacia adelante”.
11. Ana se quedó embarazada de Miguel cuando tenía 19 años.
12. La madre empezó a ponerse furiosa a causa de todo el desorden en el cuarto del niño.

- b) ¿Qué diferencia hay entre los verbos del ejercicio a)? Rellena la tabla siguiente.¹³⁴¹
 ¿Con qué clase de palabra pueden ser combinados los equivalentes del alemán *werden*?

significado	equivalente de <i>werden</i>	+	clase de palabra
Grundlegende Veränderung / Wandlung (auch Berufe)	<u>convertirse</u> en / meterse	+	<u>sustantivo</u>
Langsame Entwicklung / allmähliche Veränderung	<u>hacerse</u>	+	adjetivo / <u>sustantivo</u>
Plötzliche, überraschende, starke Gemütsveränderung	volverse	+	<u>adjetivo</u> / sustantivo
Endgültige Veränderung / Ergebnis	quedarse	+	<u>adjetivo</u>
Vorübergehender Zustand / Veränderung	<u>ponerse</u>	+	adjetivo
Absicht / Vorhaben	<u>querer ser</u>	+	adjetivo / <u>sustantivo</u>
Erreichen eines Ziels	llegarse a ser	+	adjetivo / <u>sustantivo</u>

c) También hay palabras propias para expresar el alemán *werden*.

Verwende die neue, annotierte Version des CREA, um die korrekten Verben zu finden, wie im untenstehenden Beispiel (<http://web.frl.es/CREA/view/inicioExterno.view>). Als *Lema* gibst du den Wortstamm zwischen zwei Asterisken ein (z. B.: **peor**) und wählst *verbo* im Bereich *Clase de palabra* sowie *infinitivo* im Bereich *Tiempo* aus.¹³⁴² Nach einem Klick auf das Feld *Concordancia* erscheint eine Liste mit Beispielsätzen, an denen du das korrekte Verb ablesen kannst.

Concordancias | Coapariciones | Configuración | Ayuda | Modo de cita | Sugerencias

Lema: **peor** Forma: Clase de palabra: verbo Grafía original: + Subcorpus

Modo: Tiempo: infinitivo Persona: Número: Tipo:

+ - Concordancia Estadística Nueva consulta

265 casos en 234 documentos.

REF. (Clasificación, país)	CONCORDANCIA	Ordenar por
1 1975 Méx.	Lupita: (Sentenciosa.) Siempre que algo cambia es para empeorar .	
2 1975 Esp.	Tramacho moviendo la cabeza dubitativamente: «Falsas ilusiones, amigo: no tardará en empeorar para siempre. He visto	
3 1975 Arg.	no recurran jamás a los retos, humillaciones, castigos y burlas que sólo han de empeorar la situación.	
4 1976 Ven.	podrían utilizar las drogas como tranquilizantes. Sin embargo las mismas drogas pueden empeorar su condición mental, a	
5 1976 Ven.	interpretada como empeoramiento del cuadro esquizofrénico, aumentando la dosis para empeorar todavía más los síntom	

CREA: Suche nach einem mit dem Adjektiv *peor* verwandten Verb.

schlechter werden: **peor** → empeorar

besser werden: **mejor** → mejorar

dick werden: **gord** → engordar

schlank werden: **delga** → adelgazar

krank werden: **enferm** → enfermar

traurig werden: **triste** → entristecer

Nacht werden: **noche** → anohecer

dunkel werden: **oscur** → oscurecer

- d) Ahora escribe una frase para cada verbo del ejercicio c) en tu cuaderno de ejercicios.
 Puedes utilizar el CORPES para encontrar un contexto apropiado.

¹³⁴¹ Einteilung in Anlehnung an: Görrissen (2005): op. cit., S. 220-221. Und De Bruyne (2002): op. cit., S. 578-583.

¹³⁴² Der Platzhalter *,* bedeutet *kein, ein oder mehr als ein Zeichen*. Auf diese Weise werden Wörter gesucht, die aus dem Wortstamm und einem optionalen Prä- und/oder Suffix variabler Länge bestehen.

21. Arbeitsblatt 12: ¿Ser, estar o hay? – Korpus: Sacodeyl

¿Ser, estar o hay?

Completa las frases siguientes con las formas correctas de los verbos *ser*, *estar* o *hay*.

1. Me encanta _____ con mis abuelos.
2. Aquí _____ una escuela de arte dramático.
3. ¿Te gustaría dedicarte a _____ profesor?
4. Pero yo sé que luego voy a _____ más contenta.
5. Ahora mismo _____ en segundo de bachiller.
6. También me gustaría periodismo deportivo, pero _____ muchos periodistas aquí en España y _____ menos empleo.
7. Nunca te das cuenta de que _____ bien hasta que te pasa algo malo.
8. _____ fan de 'El señor de los anillos'. ¿Has visto la película también?
9. Me gusta este barrio porque todo _____ muy cerca – el centro comercial, la biblioteca, el instituto.
10. Mis dos padres _____ profesores. Mi madre _____ profesor de francés y mi padre de química.
11. Cuando tengo poco trabajo me gusta _____ en el gimnasio.
12. Los animales _____ una de mis pasiones.
13. Hay otros países que _____ en peores situaciones que los que _____ en la Unión Europea.
14. Me gusta todo tipo de películas y luego _____ otras que no veo nunca.
15. _____ en un grupo de teatro.
16. En tu casa ¿_____ muchos, tienes muchos hermanos? – No, _____ cinco, mis padres y tres hermanos.
17. Unas pistas de skate _____ al lado del Barnés y _____ allí muchas veces.
18. También me gusta _____ aquí en mi casa en las vacaciones.
19. ¿Qué asignaturas te gustan más? – Las que más me gustan _____ matemáticas y el inglés.
20. Siempre me ha gustado el deporte, y además _____ muy sano, pero cuando no _____ tiempo y tienes que _____ estudiando, no se puede hacer nada.
21. Yo quiero _____ actriz, _____ mi sueño de toda la vida.
22. Practico fútbol, pero no _____ en ningún equipo.
23. Ahora mismo _____ estudiando cuarto de la E.S.O. – _____ cuatro cursos de la E.S.O. y dos de bachillerato.
24. Últimamente no tengo mucho tiempo porque _____ bastantes exámenes.

¿Ser, estar o hay? – Solución

Completa las frases siguientes con las formas correctas de los verbos *ser*, *estar* o *hay*.

1. Me encanta estar con mis abuelos.
2. Aquí hay una escuela de arte dramático.
3. ¿Te gustaría dedicarte a ser profesor?
4. Pero yo sé que luego voy a estar más contenta.
5. Ahora mismo estoy en segundo de bachiller.
6. También me gustaría periodismo deportivo, pero hay muchos periodistas aquí en España y hay menos empleo.
7. Nunca te das cuenta de que estás bien hasta que te pasa algo malo.
8. Soy fan de 'El señor de los anillos'. ¿Has visto la película también?
9. Me gusta este barrio porque todo está muy cerca – el centro comercial, la biblioteca, el instituto.
10. Mis dos padres son profesores. Mi madre es profesor de francés y mi padre de química.
11. Cuando tengo poco trabajo me gusta estar en el gimnasio.
12. Los animales son una de mis pasiones.
13. Hay otros países que están en peores situaciones que los que están en la Unión Europea.
14. Me gusta todo tipo de películas y luego hay otras que no veo nunca.
15. Estoy en un grupo de teatro.
16. En tu casa ¿sois muchos, tienes muchos hermanos? – No, somos cinco, mis padres y tres hermanos.
17. Unas pistas de skate están al lado del Barnés y estamos allí muchas veces.
18. También me gusta estar aquí en mi casa en las vacaciones.
19. ¿Qué asignaturas te gustan más? – Las que más me gustan son matemáticas y el inglés.
20. Siempre me ha gustado el deporte, y además es muy sano, pero cuando no hay tiempo y tienes que estar estudiando, no se puede hacer nada.
21. Yo quiero ser actriz, es mi sueño de toda la vida.
22. Practico fútbol, pero no estoy en ningún equipo.
23. Ahora mismo estoy estudiando cuarto de la E.S.O. – hay cuatro cursos de la E.S.O. y dos de bachillerato.
24. Últimamente no tengo mucho tiempo porque hay bastantes exámenes.

22. Arbeitsblatt 13: *Adjetivos con ser y estar* – Korpus: CREA**Adjetivos con *ser* y *estar***

Verwende das *Corpus de Referencia del Español Actual* (CREA), um zu ermitteln, welche Bedeutung die folgenden Adjektive¹³⁴³ mit dem jeweiligen Verb haben (<http://web.frl.es/CREA/view/inicioExterno.view>). Zunächst gibst du das Verb *ser* als *Lema* ein, klickst auf das Feld *Proximidad* und trägst das jeweilige Adjektiv (z. B.: *rico*) als *Lema* der zweiten Zeile ein. Daraufhin klickst du auf das Pluszeichen am unteren Rand der Eingabemaske (nicht das Pluszeichen in der zweiten Eingabezeile), um ein weiteres Eingabemenü aufzurufen. Hier verfährt du auf die gleiche Weise mit dem Verb *estar* und dem jeweiligen Adjektiv. Für beide Kombinationen wählst du ein Intervall von einem Wort rechts des Suchbegriffs. Nach einem Klick auf das Feld *Concordancia* erscheinen Konkordanzen mit diesem Adjektiv und einem der beiden Verben. Lies dir einige Sätze durch, und entscheide, welche Bedeutung das Adjektiv mit welchem Verb hat. Markiere die Felder der Tabelle in grün für die Bedeutung mit dem Verb *ser* oder gelb für die Bedeutung mit dem Verb *estar*. Danach wählst du diejenige Farbe für das Feld *más usado*, deren Verb häufiger zusammen mit dem Adjektiv erscheint.

CREA: Suche nach *rico* mit den Verben *ser* und *estar*.

adjetivo	significado 1	más usado	significado 2
rico	reich		lecker
listo	fertig, bereit		schlau, klug
aburrido	gelangweilt, es leid sein		langweilig, fade
delicado	feinfühlig, heikel		gebrechlich, kränklich
atento	aufmerksam = achtsam		aufmerksam = höflich
callado	still = schweigsam, ruhig		still = leise, still, schweigend
despierto	aufgeweckt		wach
orgulloso	stolz (Art, Charakter)		stolz auf jemanden / etwas
cansado	müde		ermüdend
seco	getrocknet, trocken		trocken (Charakter, Klima)

¹³⁴³ Einteilung der Adjektive aus: Görrissen (2005): op. cit., S. 208-209.

Adjektivos con *ser* y *estar* – Solución

Verwende das *Corpus de Referencia del Español Actual* (CREA), um zu ermitteln, welche Bedeutung die folgenden Adjektive¹³⁴⁴ mit dem jeweiligen Verb haben (<http://web.frl.es/CREA/view/inicioExterno.view>). Zunächst gibst du das Verb *ser* als *Lema* ein, klickst auf das Feld *Proximidad* und trägst das jeweilige Adjektiv (z. B.: *rico*) als *Lema* der zweiten Zeile ein. Daraufhin klickst du auf das Pluszeichen am unteren Rand der Eingabemaske (nicht das Pluszeichen in der zweiten Eingabezeile), um ein weiteres Eingabemenü aufzurufen. Hier verfahrst du auf die gleiche Weise mit dem Verb *estar* und dem jeweiligen Adjektiv. Für beide Kombinationen wählst du ein Intervall von einem Wort rechts des Suchbegriffs. Nach einem Klick auf das Feld *Concordancia* erscheinen Konkordanzen mit diesem Adjektiv und einem der beiden Verben. Lies dir einige Sätze durch, und entscheide, welche Bedeutung das Adjektiv mit welchem Verb hat. Markiere die Felder der Tabelle in grün für die Bedeutung mit dem Verb *ser* oder gelb für die Bedeutung mit dem Verb *estar*. Danach wählst du diejenige Farbe für das Feld *más usado*, deren Verb häufiger zusammen mit dem Adjektiv erscheint.

Corpus de Referencia del Español Actual
(CREA. Versión anotada)

Versión beta (0.1)

REAL ACADEMIA ESPAÑOLA

Concordancias | Coapariciones | Configuración | Ayuda | Modo de cita | Sugerencias

Conector

☐ Lema ser Forma Clase de palabra (Todos) Grafía original + Subcorpus - Proximidad

Proximidad Limpiar

Lema rico Forma Clase de palabra (Todos) Distancia Intervalo 1 Izquierda Derecha

☐ Lema estar Forma Clase de palabra (Todos) Grafía original + Subcorpus - Proximidad

Proximidad Limpiar

Lema rico Forma Clase de palabra (Todos) Distancia Intervalo 1 Izquierda Derecha

+ - Concordancia Estadística Nueva consulta

CREA: Suche nach *rico* mit den Verben *ser* und *estar*.

adjetivo	significado 1	más usado	significado 2
rico	reich		lecker
listo	fertig, bereit		schlau, klug
aburrido	gelangweilt, es leid sein		langweilig, fade
delicado	feinfühlig, heikel		gebrechlich, kränklich
atento	aufmerksam = achtsam		aufmerksam = höflich
callado	still = schweigsam, ruhig		still = leise, still, schweigend
despierto	aufgeweckt		wach
orgulloso	stolz (Art, Charakter)		stolz auf jemanden / etwas
cansado	müde		ermüdend
seco	getrocknet, trocken		trocken (Charakter, Klima)

¹³⁴⁴ Einteilung der Adjektive aus: Görrissen (2005): op. cit., S. 208-209.

23. Arbeitsblatt 14: *Preposiciones* – Korpus: CdE:Nuevo**Preposiciones****a) Encuentra la preposición correcta para estos verbos.**

Verwende dafür die Funktion *Colocados* des CdE:Nuevo (<https://www.corpusdelespanol.org/web-dial/>). Gib das Verb (z. B.: *preocuparse*) als *Palabra/frase* ein, wähle *prep.todos* im Bereich *Colocados* und zwei Einheiten nach rechts als Intervall aus. Nach einem Klick auf das Feld *Buscar* erscheint eine Auflistung aller Präpositionen, die dem jeweiligen Verb im Korpus folgen. Wähle diejenigen Präpositionen aus, die infrage kommen, und lass dir mittels eines Klicks auf das Feld *Contexto* die Konkordanzen anzeigen. Lies dir einige Sätze durch, und entscheide, welche Präposition die richtige für den unten angegebenen Kontext ist.

CONTEXTO	FREC
POR	81
DE	24
EN	4
A	4
COMO	3
SOBRE	2
CON	1
ANTES	1
ACERCA	1
TOTAL	121

Links: Suche nach Präpositionen mit *preocuparse*. Rechts: Präpositionen, die mit *preocuparse* erscheinen.

expresión alemana	verbo español	preposición correcta
1. sich Sorgen machen um	preocuparse	<u>por</u>
2. sich kümmern um	preocuparse	
3. sich über etwas freuen	alegrarse	
4. träumen von	soñar	
5. für etwas sein	estar a favor	
6. gegen etwas sein	estar en contra	
7. sich zu etwas anmelden	apuntarse	
8. bei etwas mitarbeiten	colaborar	
9. mit jemandem zusammenarbeiten	colaborar	
10. sich auf etwas konzentrieren	concentrarse	
11. sich verwandeln in	convertirse	
12. einwandern	inmigrar	

b) Elige seis de las combinaciones de arriba y escribe una frase con cada una.

Puedes usar los ejemplos del corpus para esta tarea. Si no conoces algunas palabras, usa el diccionario en línea *PONS* (<https://de.pons.com>).

Preposiciones – Solución

a) Encuentra la preposición correcta para estos verbos.

Verwende dafür die Funktion *Colocados* des CdE:Nuevo (<https://www.corpusdelespanol.org/web-dial/>). Gib das Verb (z. B.: *preocuparse*) als *Palabra/frase* ein, wähle *prep.todos* im Bereich *Colocados* und zwei Einheiten nach rechts als Intervall aus. Nach einem Klick auf das Feld *Buscar* erscheint eine Auflistung aller Präpositionen, die dem jeweiligen Verb im Korpus folgen. Wähle diejenigen Präpositionen aus, die infrage kommen, und lass dir mittels eines Klicks auf das Feld *Contexto* die Konkordanzen anzeigen. Lies dir einige Sätze durch, und entscheide, welche Präposition die richtige für den unten angegebenen Kontext ist.

Links: Suche nach Präpositionen mit *preocuparse*. Rechts: Präpositionen, die mit *preocuparse* erscheinen.

expresión alemana	verbo español	preposición correcta
1. sich Sorgen machen um	preocuparse	<u>por</u>
2. sich kümmern um	preocuparse	de
3. sich über etwas freuen	alegrarse	de
4. träumen von	soñar	con
5. für etwas sein	estar a favor	de
6. gegen etwas sein	estar en contra	de
7. sich zu etwas anmelden	apuntarse	a
8. bei etwas mitarbeiten	colaborar	en
9. mit jemandem zusammenarbeiten	colaborar	con
10. sich auf etwas konzentrieren	concentrarse	en
11. sich verwandeln in	convertirse	en
12. einwandern	inmigrar	a

b) Elige seis de las combinaciones de arriba y escribe una frase con cada una.

Puedes usar los ejemplos del corpus para esta tarea. Si no conoces algunas palabras, usa el diccionario en línea *PONS* (<https://de.pons.com>).

1. Angelina se alejó de su clase de actuación y comenzó a soñar con ser directora.
2. Preocuparse por el futuro es absurdo, porque el futuro está sucediendo ahora mismo.
3. Si estás en contra de algún político en particular, debes estar a favor de su oposición.
4. Marina dedica su tiempo a colaborar con numerosas organizaciones de beneficio.
5. Decían que tendrían la posibilidad de hacer carrera y de convertirse en vendedores profesionales.
6. A los 18 años decidió inmigrar a Israel.

24. Arbeitsblatt 15: *El presente: Formas irregulares* – Korpus: CORPES**El presente: Formas irregulares**

Encuentra las formas correctas de los verbos irregulares en presente.

Verwende dafür das *Corpus del Español del Siglo XXI* – CORPES (<http://web.frl.es/CORPES/view/inicioExterno.view>). Zunächst gibst du das Verb (z. B.: *conocer*) als *Lema* ein, wählst die Wortart (*verbo*), den Modus (*indicativo*) und die Zeit (*presente*) aus. Daraufhin bestimmst du die Person (z. B.: *primera*) und den Numerus (z. B.: *singular*) und klickst auf *Concordancia*.

Concordancias | Coapariciones | Configuración | Ayuda | Modo de cita | Sugerencias

Lema: Forma: Clase de palabra: ☐ Grafía original

Modo: Tiempo: Persona: Número: Tipo:

8.524 casos en 4.094 documentos.

REF. (Clasificación, país)	CONCORDANCIA	Ordenar
1 2001 Arg.	dentro de tu disco anterior (tus ídolos)? No me refiero a los músicos, sino a los otros. Conozco a algunos, pero a	
2 2001 Arg.	Charli es un genio, el único que conozco . Todo lo que hace	
3 2001 Arg.	He oído muy poco de Javier y al otro no le conozco .	
4 2001 Cuba	Perdón, no lo conozco .	
5 2001 Arg.	Si, lo conozco pero no estoy en c	
6 2001 Esp.	Conozco poco de ellos.	
7 2001 Esp.	No los conozco	

CORPES: Suche nach der Verbform der 1. Person Singular von *conocer* im *presente de indicativo*.

Sigue con la segunda y tercera persona singular y después con las formas del plural.

Repaso: Formas regulares:

Rellena la tabla con las formas regulares o de la memoria o con la ayuda del corpus.

	trabajar	beber	vivir
1. pers. sg.	trabaj- [beb- [viv- [
2. pers. sg.			
3. pers. sg.			
1. pers. pl.			
2. pers. pl.			
3. pers. pl.			

Formas irregulares:

En el presente no solo hay formas regulares sino también hay verbos irregulares. Repasa las formas de los verbos irregulares con la ayuda del corpus. Hay verbos completamente irregulares, verbos con diptongación y verbos en los que solo la primera persona singular es irregular.

1) Irregularidades en la primera persona singular:

¿Cuál es la forma correcta de la primera persona singular de estos verbos? Las otras formas son regulares.

	conocer	dar	hacer	poner
1. pers. sg.	_____	_____	_____	_____

	saber	salir	traer	ver
1. pers. sg.	_____	_____	_____	_____

2) Verbos con diptongación:

	pensar	querer	mostrar	poder	dormir
1. pers. sg.	_____	_____	_____	_____	_____
2. pers. sg.	_____	_____	_____	_____	_____
3. pers. sg.	_____	_____	_____	_____	_____
1. pers. pl.	_____	_____	_____	_____	_____
2. pers. pl.	_____	_____	_____	_____	_____
3. pers. pl.	_____	_____	_____	_____	_____
	_____	_____	_____	_____	_____

¿Cuál es la conjugación correcta de los verbos *volver*, *empezar*, *jugar* y *contar*?
Escribe estos verbos debajo de la conjugación correcta (última línea).

3) Verbos con irregularidad en la primera persona singular y con diptongación:

	venir	tener	decir
1. pers. sg.	_____	_____	_____
2. pers. sg.	_____	_____	_____
3. pers. sg.	_____	_____	_____
1. pers. pl.	_____	_____	_____
2. pers. pl.	_____	_____	_____
3. pers. pl.	_____	_____	_____

El presente: Formas irregulares – Solución

Encuentra las formas correctas de los verbos irregulares en presente.

Verwende dafür das *Corpus del Español del Siglo XXI* – CORPES (<http://web.frl.es/CORPES/view/inicioExterno.view>). Zunächst gibst du das Verb (z. B.: *conocer*) als *Lema* ein, wählst die Wortart (*verbo*), den Modus (*indicativo*) und die Zeit (*presente*) aus. Daraufhin bestimmst du die Person (z. B.: *primera*) und den Numerus (z. B.: *singular*) und klickst auf *Concordancia*.

Concordancias | Coapariciones | Configuración | Ayuda | Modo de cita | Sugerencias

Lema: Forma: Clase de palabra: ☐ Grafía original

Modo: Tiempo: Persona: Número: Tipo:

8.524 casos en 4.094 documentos.

REF. (Clasificación, país)	CONCORDANCIA	Ordenar
1 2001 Arg.	dentro de tu disco anterior (tus ídolos)? No me refiero a los músicos, sino a los otros. Conozco a algunos, pero a	
2 2001 Arg.	Charli es un genio, el único que conozco . Todo lo que hace	
3 2001 Arg.	He oído muy poco de Javier y al otro no le conozco .	
4 2001 Cuba	Perdón, no lo conozco .	
5 2001 Arg.	Sí, lo conozco pero no estoy en c	
6 2001 Esp.	Conozco poco de ellos.	
7 2001 Esp.	No los conozco	

CORPES: Suche nach der Verbform der 1. Person Singular von *conocer* im *presente de indicativo*.

Sigue con la segunda y tercera persona singular y después con las formas del plural.

Repaso: Formas regulares:

Rellena la tabla con las formas regulares o de la memoria o con la ayuda del corpus.

	trabajar	beber	vivir
1. pers. sg.	trabaj-	- <u>o</u>	- <u>o</u>
2. pers. sg.		- <u>es</u>	- <u>es</u>
3. pers. sg.		- <u>e</u>	- <u>e</u>
1. pers. pl.		- <u>emos</u>	- <u>imos</u>
2. pers. pl.		- <u>éis</u>	- <u>ís</u>
3. pers. pl.		- <u>en</u>	- <u>en</u>

Formas irregulares:

En el presente no solo hay formas regulares sino también hay verbos irregulares. Repasa las formas de los verbos irregulares con la ayuda del corpus. Hay verbos completamente irregulares, verbos con diptongación y verbos en los que solo la primera persona singular es irregular.

1) Irregularidades en la primera persona singular:

¿Cuál es la forma correcta de la primera persona singular de estos verbos? Las otras formas son regulares.

	conocer	dar	hacer	poner
1. pers. sg.	<u>conozco</u>	<u>doy</u>	<u>hago</u>	<u>pongo</u>

	saber	salir	traer	ver
1. pers. sg.	<u>sé</u>	<u>salgo</u>	<u>traigo</u>	<u>veo</u>

2) Verbos con diptongación:

	pensar	querer	mostrar	poder	dormir
1. pers. sg.	<u>pienso</u>	<u>quiero</u>	<u>muestro</u>	<u>puedo</u>	<u>duermo</u>
2. pers. sg.	<u>piensas</u>	<u>quieres</u>	<u>muestras</u>	<u>puedes</u>	<u>duermes</u>
3. pers. sg.	<u>piensa</u>	<u>quiere</u>	<u>muestra</u>	<u>puede</u>	<u>duerme</u>
1. pers. pl.	<u>pensamos</u>	<u>queremos</u>	<u>mostramos</u>	<u>podemos</u>	<u>dormimos</u>
2. pers. pl.	<u>pensáis</u>	<u>queréis</u>	<u>mostráis</u>	<u>podéis</u>	<u>dormís</u>
3. pers. pl.	<u>piensan</u>	<u>quieren</u>	<u>muestran</u>	<u>pueden</u>	<u>duermen</u>
	<u>empezar</u>		<u>contar, jugar</u>	<u>volver</u>	

¿Cuál es la conjugación correcta de los verbos *volver*, *empezar*, *jugar* y *contar*?
Escribe estos verbos debajo de la conjugación correcta (última línea).

3) Verbos con irregularidad en la primera persona singular y con diptongación:

	venir	tener	decir
1. pers. sg.	<u>vengo</u>	<u>tengo</u>	<u>digo</u>
2. pers. sg.	<u>vienes</u>	<u>tienes</u>	<u>dices</u>
3. pers. sg.	<u>viene</u>	<u>tiene</u>	<u>dice</u>
1. pers. pl.	<u>venimos</u>	<u>tenemos</u>	<u>decimos</u>
2. pers. pl.	<u>venís</u>	<u>tenéis</u>	<u>decís</u>
3. pers. pl.	<u>vienen</u>	<u>tienen</u>	<u>dicen</u>

25. Arbeitsblatt 16: *El pretérito indefinido: Formas* – Korpus: CORPES**El pretérito indefinido: Formas****Encuentra las formas correctas del pretérito indefinido.**

Verwende dafür das *Corpus del Español del Siglo XXI* – CORPES (<http://web.frl.es/CORPES/view/inicioExterno.view>). Zunächst gibst du das Verb (z. B.: *continuar*) als *Lema* ein und wählst die Wortart (*verbo*) sowie die Zeit (*pasado*) aus. Daraufhin bestimmst du Person (z. B.: *primera*) und Numerus (z. B.: *singular*) und klickst auf *Concordancia*.

Concordancias | Coapariciones | Configuración | Ayuda | Modo de cita | Sugerencias

Lema: Forma: Clase de palabra: ☐ Grafía original

Modo: Tiempo: Persona: Número: Tipo:

619 casos en 370 documentos.

REF. (Clasificación, país)	CONCORDANCIA	Ordenar
1 2001 Col.	incógnita de su adiós entre las palmeras cotidianas. Sin mover mis ojos hacia el llanto, continué saboreando un ca	
2 2001 Ur.	«Está con uno de sus amantes», pensé. Había perdido el control y un poco después continué la destrucción en e	
3 2001 Chile	inquietante, iba a abrir las puertas del pasado pero resistí, no era el momento. Continué por inercia, por gus	
4 2001 Chile	sentarme. La pared dejaba un hueco de medio metro, todo un mundo. Sin descansar, continué . La montaña segui	
5 2001 Hond.	arrastrando la pronunciación como si fuera un descubrimiento digno de pregonar, y continué hablándole morda	
6 2001 Hond.	scotch con una mujer de verdad, así que decidí relajarme y comportarme como soy, y continué :	
7 2001 Hond.	pensar, luego me agarre el mentón, recorrí su cuerpo de pies a ojos con la mirada y continué , en tono magistral:	
8 2001 Hond.	micrófono frente a mi cara, como intimidándome en respuesta a mis miradas lesivas, pero continué con el discurso:	
9 2001 Hond.	tan extravagantes que ningún subconsciente podría soportar tenerlos reprimidos - continué lentamente, levant	
10 2001 Hond.	pero cuando ella se percató de mi pausa contemplativa y se animó a decir algo, yo continué malcriadamente:	
11 2001 Hond.	una aparatosa victoria de los instintos sobreponiéndose a la sub cultura, así que continué :	
12 2001 Hond.	-Bienvenida al mundo de las cantinas bohemias - continué , improvisando de e	
13 2001 Méx.	aves hembras incapaces de resistir el reclamo de su macho. Con brinco en la panza continué con la espera que :	
14 2001 Méx.	, dejándome en plena conciencia del momento que vivía. Sigue, ordenó El Negro. Y continué , ya sin ti. Ahora ac	
15 2001 Méx.	indignidades de cada noche en Las Guanábanas). Esta vez tropezó con mi callada rebeldía: continué amamantando a m	

CORPES: Suche nach der Verbform der 1. Person Singular von *continuar* im *pretérito indefinido*.

Sigue con la segunda y tercera persona singular y después con las formas del plural.

Formas regulares:

	continuar	beber	vivir
1. pers. sg.	continué	_____	_____
2. pers. sg.	_____	_____	_____
3. pers. sg.	_____	_____	_____
1. pers. pl.	_____	_____	_____
2. pers. pl.	_____	_____	_____
3. pers. pl.	_____	_____	_____

Formas irregulares:

infinitivo	pretérito indefinido	
<div> <div> <div>estar</div> <div>querer</div> <div>poder</div> <div>poner</div> <div>tener</div> <div>saber</div> <div>venir</div> </div> <div> <div>_____</div> <div>_____</div> <div>_____</div> <div>_____</div> <div>_____</div> <div>_____</div> <div>_____</div> </div> </div>		<div> <div> <div>-e</div> <div>-iste</div> <div>-o</div> <div>-imos</div> <div>-isteis</div> <div>-ieron</div> </div> </div>

	hacer	dar	ver
1. pers. sg.	_____	_____	_____
2. pers. sg.	_____	_____	_____
3. pers. sg.	_____	_____	_____
1. pers. pl.	_____	_____	_____
2. pers. pl.	_____	_____	_____
3. pers. pl.	_____	_____	_____

	ser	ir	dormir	llegar
1. pers. sg.	_____	_____	_____	_____
2. pers. sg.	_____	_____	_____	_____
3. pers. sg.	_____	_____	_____	_____
1. pers. pl.	_____	_____	_____	_____
2. pers. pl.	_____	_____	_____	_____
3. pers. pl.	_____	_____	_____	_____

El pretérito indefinido: Formas

Encuentra las formas correctas del pretérito indefinido.

Verwende dafür das *Corpus del Español del Siglo XXI* – CORPES (<http://web.frl.es/CORPES/view/inicioExterno.view>). Zunächst gibst du das Verb (z. B.: *continuar*) als *Lema* ein und wählst die Wortart (*verbo*) sowie die Zeit (*pasado*) aus. Daraufhin bestimmst du Person (z. B.: *primera*) und Numerus (z. B.: *singular*) und klickst auf *Concordancia*.

Concordancias | Coapariciones | Configuración | Ayuda | Modo de cita | Sugerencias

Lema: Forma: Clase de palabra: ☐ Grafía original

Modo: Tiempo: Persona: Número: Tipo:

619 casos en 370 documentos.

REF. (Clasificación, país)	CONCORDANCIA	Ordenar
1 2001 Col.	incógnita de su adiós entre las palmeras cotidianas. Sin mover mis ojos hacia el llanto, continué saboreando un ca	
2 2001 Ur.	«Está con uno de sus amantes», pensé. Había perdido el control y un poco después continué la destrucción en e	
3 2001 Chile	inquietante, iba a abrir las puertas del pasado pero resistí, no era el momento. Continué por inercia, por gue	
4 2001 Chile	sentarme. La pared dejaba un hueco de medio metro, todo un mundo. Sin descansar, continué . La montaña segui	
5 2001 Hond.	arrastrando la pronunciación como si fuera un descubrimiento digno de pregonar, y continué hablándole morda	
6 2001 Hond.	scotch con una mujer de verdad, así que decidí relajarme y comportarme como soy, y continué :	
7 2001 Hond.	pensar, luego me agarre el mentón, recorrí su cuerpo de pies a ojos con la mirada y continué , en tono magistral:	
8 2001 Hond.	micrófono frente a mi cara, como intimidándome en respuesta a mis miradas lesivas, pero continué con el discurso:	
9 2001 Hond.	tan extravagantes que ningún subconsciente podría soportar tenerlos reprimidos - continué lentamente, levant	
10 2001 Hond.	pero cuando ella se percató de mi pausa contemplativa y se animó a decir algo, yo continué malcriadamente:	
11 2001 Hond.	una aparatosa victoria de los instintos sobreponiéndose a la sub cultura, así que continué :	
12 2001 Hond.	-Bienvenida al mundo de las cantinas bohemias - continué , improvisando de e	
13 2001 Méx.	aves hembras incapaces de resistir el reclamo de su macho. Con bríncos en la panza continué con la espera que :	
14 2001 Méx.	, dejándome en plena conciencia del momento que vivía. Sigue, ordenó El Negro. Y continué , ya sin ti. Ahora ac	
15 2001 Méx.	indignidades de cada noche en Las Guanábanas). Esta vez tropezó con mi callada rebeldía: continué amamantando a m	

CORPES: Suche nach der Verbform der 1. Person Singular von *continuar* im *pretérito indefinido*.

Sigue con la segunda y tercera persona singular y después con las formas del plural.

Formas regulares:

	continuar	beber	vivir
1. pers. sg.	continué	bebí	viví
2. pers. sg.	continuaste	bebiste	viviste
3. pers. sg.	continuó	bebió	vivió
1. pers. pl.	continuamos	bebimos	vivimos
2. pers. pl.	continuasteis	bebisteis	vivisteis
3. pers. pl.	continuaron	bebieron	vivieron

Formas irregulares:

infinitivo	pretérito indefinido	
estar	<u>estuv-</u>	[-e -iste -o -imos -isteis -ieron
querer	<u>quis-</u>	
poder	<u>pud-</u>	
poner	<u>pus-</u>	
tener	<u>tuv-</u>	
saber	<u>sup-</u>	
venir	<u>vin-</u>	

	hacer	dar	ver
1. pers. sg.	<u>hice</u>	<u>di</u>	<u>vi</u>
2. pers. sg.	<u>hiciste</u>	<u>diste</u>	<u>viste</u>
3. pers. sg.	<u>hizo</u>	<u>dio</u>	<u>vio</u>
1. pers. pl.	<u>hicimos</u>	<u>dimos</u>	<u>vimos</u>
2. pers. pl.	<u>hicisteis</u>	<u>disteis</u>	<u>visteis</u>
3. pers. pl.	<u>hicieron</u>	<u>dieron</u>	<u>vieron</u>

	ser	ir	dormir	llegar
1. pers. sg.	<u>fui</u>	<u>fui</u>	<u>dormí</u>	<u>llegué</u>
2. pers. sg.	<u>fuiste</u>	<u>fuiste</u>	<u>dormiste</u>	<u>llegaste</u>
3. pers. sg.	<u>fue</u>	<u>fue</u>	<u>durmió</u>	<u>llegó</u>
1. pers. pl.	<u>fuimos</u>	<u>fuimos</u>	<u>dormimos</u>	<u>llegamos</u>
2. pers. pl.	<u>fuisteis</u>	<u>fuisteis</u>	<u>dormisteis</u>	<u>llegasteis</u>
3. pers. pl.	<u>fueron</u>	<u>fueron</u>	<u>durmieron</u>	<u>llegaron</u>

26. Arbeitsblatt 17: *El participio de pasado: Formas regulares e irregulares* – Korpus: CORPES

El participio de pasado: Formas regulares e irregulares

Verwende das *Corpus del Español del Siglo XXI* – CORPES (<http://web.frl.es/CORPES/view/inicioExterno.view>). Zunächst gibst du das Verb (z. B.: *hablar*) als *Lema* ein und wählst die Wortart (*verbo*) sowie die Zeit (*participio de pasado*) aus. Nach einem Klick auf das Feld *Concordancia* erscheint eine Liste mit Beispielsätzen mit der jeweiligen Partizipialform, welche du in die Tabelle eintragen kannst.

Concordancias | Coapariciones | Configuración | Ayuda | Modo de cita | Sugerencias

Lema: Forma: Clase de palabra: ☐ Grafía original

Modo: Tiempo: Persona: Número: Tipo:

8.312 casos en 4.420 documentos.

REF. (Clasificación, país)	CONCORDANCIA	Ordenar
1 2001 Méx.	No sé si te he hablado de mi predecesor.	
2 2001 Chile	papelera. Duermes y tiene pesadillas, pero cuando despierta está seguro de que no ha hablado en sueños. Al día.	
3 2001 Chile	siempre lo hubiera sido y sólo esa noche, precisamente la noche en que habíamos hablado de la muerte de la	
4 2001 Chile	cualquier cosa que me dijera iba a quedar entre nosotros dos. Dijo que siempre había hablado con franqueza cor	
5 2001 Ven.	class venezolana de aquellos años. Hoy, en pésimo estado de conservación, se ha hablado de un ambicioso p	
6 2001 Col.	provocativo y hasta mal intencionado; cabello de enredadera difícil e intrincada; hablado en un dialecto de c	
7 2001 Col.	de mi casa hasta este encierro de muerte. Tampoco sé quien me tiene ya que no he hablado con nadie, por que	
8 2001 Col.	Ha hablado de su marginalidad	

CORPES: Suche nach dem Partizip Perfekt von *hablar*.

1) Encuentra los participios de pasado de los verbos siguientes.

infinitivo	participio	infinitivo	participio	infinitivo	participio
hablar	hablado	volver		freír	
decir		hacer		dormir	
poner		escribir		acostar	
beber		salir		sentar	
pensar		dar		imprimir	
abrir		vivir		traer	
tener		llegar		devolver	
ver		haber		descubrir	
discutir		resolver		comer	
morir		romper		sentir	

2) ¿Cuáles son regulares y cuáles se forman de manera irregular? Pinta las casillas con las formas irregulares de color naranja.

El participio de pasado: Formas regulares e irregulares – Solución

Verwende das *Corpus del Español del Siglo XXI – CORPES* (<http://web.frl.es/CORPES/view/inicioExterno.view>). Zunächst gibst du das Verb (z. B.: *hablar*) als *Lema* ein und wählst die Wortart (*verbo*) sowie die Zeit (*participio de pasado*) aus. Nach einem Klick auf das Feld *Concordancia* erscheint eine Liste mit Beispielsätzen mit der jeweiligen Partizipialform, welche du in die Tabelle eintragen kannst.

Concordancias | Coapariciones | Configuración | Ayuda | Modo de cita | Sugerencias

Lema: Forma: Clase de palabra: ☐ Grafía original

Modo: Tiempo: Persona: Número: Tipo:

8.312 casos en 4.420 documentos.

REF.	(Clasificación, país)	CONCORDANCIA	Ordenar
1	2001 Méx.	No sé si te he hablado de mi predecesor.	
2	2001 Chile	papelera. Duermes y tiene pesadillas, pero cuando despierta está seguro de que no ha hablado en sueños. Al día.	
3	2001 Chile	siempre lo hubiera sido y sólo esa noche, precisamente la noche en que habíamos hablado de la muerte de la	
4	2001 Chile	cualquier cosa que me dijera iba a quedar entre nosotros dos. Dijo que siempre había hablado con franqueza con	
5	2001 Ven.	class venezolana de aquellos años. Hoy, en pésimo estado de conservación, se ha hablado de un ambicioso p	
6	2001 Col.	provocativo y hasta mal intencionado; cabello de enredadera difícil e intrincada; hablado en un dialecto de c	
7	2001 Col.	de mi casa hasta este encierro de muerte. Tampoco sé quien me tiene ya que no he hablado con nadie, por que	
8	2001 Col.	Ha hablado de su marginalidad	

CORPES: Suche nach dem Partizip Perfekt von *hablar*.

1) Encuentra los participios de pasado de los verbos siguientes.

infinitivo	participio	infinitivo	participio	infinitivo	participio
hablar	<u>hablado</u>	volver	<u>vuelto</u>	freír	<u>frito / freído</u>
decir	<u>dicho</u>	hacer	<u>hecho</u>	dormir	<u>dormido</u>
poner	<u>puesto</u>	escribir	<u>escrito</u>	acostar	<u>acostado</u>
beber	<u>bebido</u>	salir	<u>salido</u>	sentar	<u>sentado</u>
pensar	<u>pensado</u>	dar	<u>dado</u>	imprimir	<u>impreso / imprimido</u>
abrir	<u>abierto</u>	vivir	<u>vivido</u>	traer	<u>traído</u>
tener	<u>tenido</u>	llegar	<u>llegado</u>	devolver	<u>devuelto</u>
ver	<u>visto</u>	haber	<u>habido</u>	descubrir	<u>descubierto</u>
discutir	<u>discutido</u>	resolver	<u>resuelto</u>	comer	<u>comido</u>
morir	<u>muerto</u>	romper	<u>roto</u>	sentir	<u>sentido</u>

2) ¿Cuáles son regulares y cuáles se forman de manera irregular? Pinta las casillas con las formas irregulares de color naranja.

Corrección de errores – Solución

Lee las frases siguientes, subraya los errores y corrígelos a la derecha.

1. El camarero dice que no hay <u>plazas</u> libres. Tenemos que buscar otro restaurante.	sitios, mesas
2. Español es <u>un lenguaje</u> interesante.	una lengua / un idioma
3. En el año 2005 <u>he estudiado</u> unos meses en España.	estudié
4. <u>En el primero</u> imagen hay una hamburguesa.	en la primera
5. No ayudar a otra persona es <u>más malo</u> que cometer un error.	peor
6. El camarero <u>recomenda</u> el pollo asado con patatas fritas.	recomienda
7. La pizzería está <u>cerrado</u> .	cerrada
8. <u>A la nueve</u> voy a la cafetería.	a las nueve
9. <u>Después el</u> último viaje, tenía dolores de cabeza.	después del
10. Español es <u>una importante idioma</u> para todo el mundo.	un idioma importante
11. Aquí tenemos <u>un foto</u> de una paella de mariscos y de un limón.	una foto
12. En las vacaciones viajamos a Italia y <u>un otro país</u> .	otro país
13. ¿Vamos al cine? – <u>Depende</u> qué película quieres ver.	depende de
14. <u>Los dos</u> personas van a un restaurante.	las dos
15. <u>El carne</u> no me gusta, prefiero la verdura.	la carne
16. Durante el día como <u>un parte</u> de pizza.	un trozo / una porción
17. El hombre quiere beber una copa de <u>vino rojo</u> .	vino tinto
18. Creo que mis padres comen <u>más sano como yo</u> .	más sano que yo
19. <u>Al primero</u> vamos al cine y después a una pizzería.	primero / en primer lugar
20. Voy a la piscina <u>uno</u> o dos veces a la semana.	una
21. Si vas <u>a tapas</u> , hay muchas cosas fritas.	de tapas
22. Creo que nuestra generación es un poquito más <u>vago</u> que la generación de nuestros padres.	vaga
23. Salimos mucho y nos gustan fiestas en casas <u>privados</u> .	privadas
24. Estoy pensando <u>a</u> mi novia.	en
25. Cambian <u>el</u> lugar y van a otro restaurante.	de

Regla:

La letra «**c**» se pronuncia como [**k**] delante de las vocales «_____» y como [**θ/s**] delante de las vocales «_____».

La letra «**g**» se pronuncia como [**g**] delante de las vocales «_____» y como [**x**] delante de las vocales «_____».

2. Tacha las «u» que no se pronuncian.

quince	guerra	cuando	vergüenza
asegurar	guantes	querido	ciudad
antigüedad	igualdad	guerrilla	llueve
estatuas	antiguo	conseguir	juegue

Regla:

La letra «**u**» **no** se pronuncia como parte de la combinación «_____» (pronunciación como [**k**]) y además entre la letra «**g**» y las vocales «_____» (pronunciación de la letra «**g**» como [**g**]).

➔ Uso de la letra «**u**» para evitar la pronunciación de la letra «**g**» como [**x**].

Si en una palabra se pronuncia la letra «**u**» entre la letra «**g**» y uno de las vocales «_____», la letra «**u**» debe llevar puntos arriba «**ü**» (este signo se llama diéresis).

3. Elige la pronunciación correcta:

conexión	[x]	[ks/s]	lluvia	[j]	[l]
xenofobia	[x]	[ks/s]	aprovechar	[tʃ]	[x]
chico	[tʃ]	[x]	examen	[x]	[ks/s]
llamar:	[j]	[l]	cuchara	[tʃ]	[x]
paella:	[j]	[l]	botella	[j]	[l]
México	[x]	[ks/s]	extrañar	[x]	[ks/s]

Regla:

La letra «**x**» siempre se pronuncia como [_____] a excepción de la palabra «_____» donde se pronuncia como [_____] (razones históricas).

La combinación de letras «**ch**» siempre se pronuncia como [_____] y la combinación de letras «**ll**» siempre se pronuncia como [_____].

Regla:

La letra «**c**» se pronuncia como [k] delante de las vocales «**a, o, u**» y como [θ/s] delante de las vocales «**e, i**».

La letra «**g**» se pronuncia como [g] delante de las vocales «**a, o, u**» y como [x] delante de las vocales «**e, i**».

2. Tacha las «u» que no se pronuncian.

quince	guerra	cuando	vergüenza
asegurar	guantes	querido	ciudad
antigüedad	igualdad	guerrilla	llueve
estatuas	antiguo	conseguir	juegue

Regla:

La letra «**u**» **no** se pronuncia como parte de la combinación «**qu**» (pronunciación como [k]) y además entre la letra «**g**» y las vocales «**e, i**» (pronunciación de la letra «**g**» como [g]).

➔ Uso de la letra «**u**» para evitar la pronunciación de la letra «**g**» como [x].

Si en una palabra se pronuncia la letra «**u**» entre la letra «**g**» y uno de las vocales «**e, i**», la letra «**u**» debe llevar puntos arriba «**ü**» (este signo se llama diéresis).

3. Elige la pronunciación correcta:

conexión	{x}	[ks/s]	lluvia	[j]	{t}
xenofobia	{x}	[ks/s]	aprovechar	[t]	{x}
chico	[t]	{x}	examen	{x}	[ks/s]
llamar:	[j]	{t}	cuchara	[t]	{x}
paella:	[j]	{t}	botella	[j]	{t}
México	[x]	{ks/s}	extrañar	{x}	[ks/s]

Regla:

La letra «**x**» siempre se pronuncia como [ks/s] a excepción de la palabra «**México**» donde se pronuncia como [x] (razones históricas).

La combinación de letras «**ch**» siempre se pronuncia como [t] y la combinación de letras «**ll**» siempre se pronuncia como [j].

29. Arbeitsblatt 20: *Comprensión auditiva: El cuento de las hermanas* – Korpus: SPLLOC

Comprensión auditiva: El cuento de las hermanas

Lee los ejercicios siguientes, escucha el texto y responde a las preguntas.

1. ¿Verdadero (V) o falso (F)?

Marca con una cruz (X) la respuesta correcta y corrige las frases falsas.

	V	F
a) Sarah y Gwen visitaron Madrid y Barcelona.		
b) Se interesaron por tres monumentos en Madrid.		
c) Dentro de Santiago Bernabéu hay un museo con las copas del Real Madrid.		
d) Gwen se compró una camiseta del Real Madrid.		
e) Después comieron tapas de jamón serrano y de queso.		
f) Cogieron el tren para ir a Barcelona.		

2. Marca con una cruz (X) la respuesta o las respuestas correctas:

- | | |
|--|--|
| a) Gwen de niña cada fin de semana... | b) Durante la semana Gwen... |
| <input type="checkbox"/> ... leía un libro. | <input type="checkbox"/> ... se despertaba muy temprano. |
| <input type="checkbox"/> ... pintaba un cuadro. | <input type="checkbox"/> ... tomaba el autobús para ir al colegio. |
| <input type="checkbox"/> ... tocaba la flauta. | <input type="checkbox"/> ... hacía los deberes pronto cuando llegaba a casa. |
| <input type="checkbox"/> ... escribía un cuento. | <input type="checkbox"/> ... veía la tele por la noche. |
| c) Sarah de niña los fines de semana... | d) Durante la semana... |
| <input type="checkbox"/> ... jugaba al voleibol. | <input type="checkbox"/> ... iba al colegio en bicicleta. |
| <input type="checkbox"/> ... tocaba la guitarra. | <input type="checkbox"/> ... casi siempre llegaba tarde a clase. |
| <input type="checkbox"/> ... veía películas. | <input type="checkbox"/> ... llamaba a sus amigos para quedarse con ellos. |
| <input type="checkbox"/> ... salía con sus amigos. | <input type="checkbox"/> ... se acostaba por la tarde. |

3. Completa el texto siguiente con las palabras que faltan:

Escuchaban un ruido en el tren y después hubo un gran revuelo. Ellas creyeron que había un _____ y de repente Gwen empezó a sentir que le caía agua en _____. Sarah mientras se reía, pero después necesitó ayuda del _____. Entonces la pidió y el revisor les ayudó a _____. Estando en un asiento _____ y mejor, se tranquilizaron. Después de llegar a Barcelona comieron _____ y se rieron recordando el viaje que habían hecho.

Comprensión auditiva: El cuento de las hermanas – *Solución*

Lee los ejercicios siguientes, escucha el texto y responde a las preguntas.

1. ¿Verdadero (V) o falso (F)?

Marca con una cruz (X) la respuesta correcta y corrige las frases falsas.	V	F
a) Sarah y Gwen visitaron Madrid y Barcelona.	X	
b) Se interesaron por tres monumentos en Madrid.		X
c) Dentro de Santiago Bernabéu hay un museo con las copas del Real Madrid.	X	
d) Gwen se compró una camiseta del Real Madrid.		X
e) Después comieron tapas de jamón serrano y de queso.		X
f) Cogieron el tren para ir a Barcelona.	X	

b) Se interesaron por dos monumentos.

d) Sarah se compró una camiseta del equipo Real Madrid.

e) Solo comieron tapas de jamón serrano, no comieron tapas de queso.

2. Marca con una cruz (X) la respuesta o las respuestas correctas:

- | | |
|---|---|
| a) Gwen de niña cada fin de semana... | b) Durante la semana Gwen... |
| <input checked="" type="checkbox"/> ... leía un libro. | <input checked="" type="checkbox"/> ... se despertaba muy temprano. |
| <input checked="" type="checkbox"/> ... pintaba un cuadro. | <input type="checkbox"/> ... tomaba el autobús para ir al colegio. |
| <input type="checkbox"/> ... tocaba la flauta. | <input checked="" type="checkbox"/> ... hacía los deberes pronto cuando llegaba a casa. |
| <input checked="" type="checkbox"/> ... escribía un cuento. | <input type="checkbox"/> ... veía la tele por la noche. |
| c) Sarah de niña los fines de semana... | d) Durante la semana... |
| <input type="checkbox"/> ... jugaba al voleibol. | <input checked="" type="checkbox"/> ... iba al colegio en bicicleta. |
| <input type="checkbox"/> ... tocaba la guitarra. | <input checked="" type="checkbox"/> ... casi siempre llegaba tarde a clase. |
| <input checked="" type="checkbox"/> ... veía películas. | <input type="checkbox"/> ... llamaba a sus amigos para quedarse con ellos. |
| <input type="checkbox"/> ... salía con sus amigos. | <input type="checkbox"/> ... se acostaba por la tarde. |

3. Completa el texto siguiente con las palabras que faltan:

Escuchaban un ruido en el tren y después hubo un gran revuelo. Ellas creyeron que había un **problema** y de repente Gwen empezó a sentir que le caía agua en **la cabeza**. Sarah mientras se reía, pero después necesitó ayuda del **revisor**. Entonces la pidió y el revisor les ayudó a **cambiarse de asiento**. Estando en un asiento **seco** y mejor, se tranquilizaron. Después de llegar a Barcelona comieron **una pizza** y se rieron recordando el viaje que habían hecho.

Transcripción del texto [3:16]¹³⁴⁵

Las vacaciones de Sarah y Gwen en España, verano de dos mil seis. Sarah y Gwen decidieron visitar España y eligieron dos ciudades. Una era Madrid y la otra era Barcelona. En Madrid visitaron la ciudad y entre sus monumentos eligieron dos que les llamaron la atención y el que más les gustó fue el Santiago Bernabeu donde juega el equipo del Real Madrid uno de los mejores equipos del mundo. Y dentro de Santiago Bernabeu hay un museo con las copas y todo relacionado con el equipo y eso les gustó mucho. Incluso Sarah, que era una forofa del fútbol, se compró una camiseta de allí.

Después en Madrid también comieron tapas de jamón serrano porque es muy típico de España y bebieron vino tinto. Después de visitar Madrid decidieron ir a Barcelona y para ello cogieron el tren. Dentro del tren estaban hablando sobre su niñez y recordaban que de pequeñas ellas eran muy diferentes.

Gwen de niña cada fin de semana leía un libro, pintaba un cuadro, escribía un cuento y durante la semana ella se despertaba todos los días muy temprano. Iba al colegio, le mandaban deberes y cuando llegaba a casa hacía los deberes pronto. Y así los terminaba temprano para poder acostarse tarde o para poder leer y hacer las cosas que hacía el fin de semana.

Por el otro lado Sarah de niña, los fines de semana, jugaba al fútbol, veía películas con sus amigos y durante la semana iba al colegio en bicicleta. Y como siempre llegaba tarde casi siempre a clase. Entonces, luego no prestaba mucha atención a las cosas que decían en clase y hacía sus deberes tarde por la noche. Como no se había enterado en el colegio, tenía que llamar a sus amigos para que le dijeran las tareas que había. Entonces, ella terminaba los deberes muy tarde. Entonces se acostaba tarde.

Después de estar recordando su niñez de repente escuchaban en el tren un ruido y después de este ruido hubo un gran revuelo. Ellas creyeron que había un problema y de repente Gwen empezó a sentir que le caía agua en la cabeza y Sarah mientras se reía, pero después necesitó ayuda del revisor. Entonces se levantó y la pidió y el revisor con mucho gusto les ayudó a cambiarse de asiento. Ellas después, estando en un asiento seco y mejor, se tranquilizaron. Y después de llegar a Barcelona comieron una pizza y se rieron recordando el viaje que habían hecho.

¹³⁴⁵ Vgl. Domínguez/Mitchell/Myles/Marsden: *SPLLOC. Hermanas story. Transcript H35NTV13N.cha*. Op. cit. (11.01.2018).

30. Arbeitsblatt 21: *Comprensión auditiva: Sevilla – Un destino muy atractivo* –
Korpus: Backbone Spanish

Comprensión auditiva: Sevilla – Un destino muy atractivo

Lee los ejercicios siguientes, escucha el texto y responde a las preguntas.

1. ¿Verdadero (V) o falso (F)?

Marca con una cruz (X) la respuesta correcta y corrige las frases falsas.

	V	F
a) Rocío tiene 25 años.		
b) Rocío vive en Sevilla.		
c) Dice que Sevilla y Murcia están al sureste de España.		
d) En Sevilla hay mucha ganadería, sobre todo el cerdo y la vaca.		
e) Además hay mucha agricultura en Sevilla y se exporta mucha fruta.		
f) Del cerdo se exporta sobre todo el jamón.		

2. Contesta las preguntas siguientes en frases completas:

a) ¿Cómo es el tiempo en Sevilla?

b) ¿Qué monumentos hay en Sevilla? Nombra dos monumentos.

c) ¿Cómo es la gente de Sevilla?

3. Marca con una cruz (X) la respuesta o las respuestas correctas:

a) La Semana Santa y la Feria...

- ☐ ... son festividades típicas.
- ☐ ... son fiestas religiosas.
- ☐ ... tienen lugar en marzo.
- ☐ ... duran una semana.

b) En la Feria...

- ☐ ... las mujeres llevan trajes típicos.
- ☐ ... las mujeres se ponen flores en el pelo.
- ☐ ... se canta.
- ☐ ... se baila el flamenco.

8. Anhang

c) En la Feria la gente come...

- ☐ ... rebujitos.
- ☐ ... patatas fritas.
- ☐ ... queso.
- ☐ ... jamón.

d) En la Semana Santa...

- ☐ ... hay imágenes del Cristo y de la Virgen.
- ☐ ... llevan las imágenes por las calles de Murcia.
- ☐ ... la gente sale para rezar a la Virgen.
- ☐ ... la gente pide cosas a los nazarenos.

4. Completa lo que dice Rocío sobre el turismo en Andalucía:

Andalucía es muy atractivo turísticamente porque tiene playas _____ y en la zona de la playa hay muchas _____ para los jóvenes. Además, hace muy buen _____. A los turistas les gusta también que se puede _____ en las montañas y que se puede hacer _____ en la playa.

Comprensión auditiva: Sevilla – Un destino muy atractivo – Solución

Lee los ejercicios siguientes, escucha el texto y responde a las preguntas.

1. ¿Verdadero (V) o falso (F)?

Marca con una cruz (X) la respuesta correcta y corrige las frases falsas.	V	F
a) Rocío tiene 25 años.	X	
b) Rocío vive en Sevilla.		X
c) Dice que Sevilla y Murcia están al sureste de España.		X
d) En Sevilla hay mucha ganadería, sobre todo el cerdo y la vaca.		X
e) Además hay mucha agricultura en Sevilla y se exporta mucha fruta.	X	
f) Del cerdo se exporta sobre todo el jamón.	X	

- b) Rocío es de Sevilla, pero vive en Murcia.
 c) Dice que Murcia está al sureste y Sevilla al sur de España.
 d) Rocío solo habla del cerdo, no menciona la vaca.

2. Contesta las preguntas siguientes en frases completas:

- a) ¿Cómo es el tiempo en Sevilla?
Hace mucho sol y llueve poco.
 b) ¿Qué monumentos hay en Sevilla? Nombra dos monumentos.
En Sevilla hay la Giralda, la Torre del Oro. (la catedral, el Río del Guadalquivir)
 c) ¿Cómo es la gente de Sevilla?
La gente es muy abierta y les gusta hablar mucho.

3. Marca con una cruz (X) la respuesta o las respuestas correctas:

- | | |
|---|---|
| a) La Semana Santa y la Feria... | b) En la Feria... |
| <input checked="" type="checkbox"/> ... son festividades típicas. | <input checked="" type="checkbox"/> ... las mujeres llevan trajes típicos. |
| <input type="checkbox"/> ... son fiestas religiosas. | <input checked="" type="checkbox"/> ... las mujeres se ponen flores en el pelo. |
| <input type="checkbox"/> ... tienen lugar en marzo. | <input checked="" type="checkbox"/> ... se canta. |
| <input checked="" type="checkbox"/> ... duran una semana. | <input type="checkbox"/> ... se baila el flamenco. |
| c) En la Feria la gente come... | d) En la Semana Santa... |
| <input type="checkbox"/> ... rebujitos. | <input checked="" type="checkbox"/> ... hay imágenes del Cristo y de la Virgen. |
| <input type="checkbox"/> ... patatas fritas. | <input type="checkbox"/> ... llevan las imágenes por las calles de Murcia. |
| <input type="checkbox"/> ... queso. | <input checked="" type="checkbox"/> ... la gente sale para rezar a la Virgen. |
| <input checked="" type="checkbox"/> ... jamón. | <input type="checkbox"/> ... la gente pide cosas a los nazarenos. |

4. Completa lo que dice Rocío sobre el turismo en Andalucía:

Andalucía es muy atractivo turísticamente porque tiene playas grandes y en la zona de la playa hay muchas discotecas para los jóvenes. Además, hace muy buen clima. A los turistas les gusta también que se puede esquiar en las montañas y que se puede hacer surf en la playa.

Transcripción del texto [3:52]¹³⁴⁶

María: ¿Puedes presentarte y decirnos un poco quién eres, de dónde eres...?

Rocío: Hola, me llamo Rocío, tengo 25 años, soy de Sevilla aunque vivo en Murcia. Sevilla, que está al sur de España, y ahora vivo en Murcia, que está al sureste.

María: ¿Y puedes hablar un poco de Sevilla? En la provincia, ¿de qué se vive? ¿Qué actividad económica es la más importante allí?

Rocío: En Sevilla hay muchas industrias, hay muchos polígonos industriales, hay mucha ganadería, sobre todo el animal del cerdo. Hay también mucha agricultura, muchas verduras, mucha fruta, se exporta mucha fruta al igual que mucha carne de cerdo.

María: ¿Y qué productos en concreto se exportan por ejemplo del cerdo, qué obtenéis?

Rocío: Del cerdo, el jamón — está muy rico el jamón sobre todo.

María: Y desde el punto de vista turístico, ¿qué destacarías de Sevilla?

Rocío: Sevilla es muy turística porque tiene muy buen clima, hace mucho sol, llueve muy poco. Es una ciudad muy grande, tiene muchos monumentos, como la Giralda, la Torre del Oro, la catedral. Hay un río muy grande, que se llama Río del Guadalquivir. Y a los turistas les gusta mucho, hay muy buena comida mucha — la gente es muy abierta, les gusta hablar mucho, estar en la calle. Hay muchos comercios, tiene un centro la ciudad muy bonito.

María: Y, por ejemplo, cuéntanos — en cuanto a festividades típicas por allí ¿qué hay que sea muy típico?

Rocío: Allí hay — las fiestas típicas primero la Semana Santa, sobre el mes de marzo. Es una semana que estamos en fiestas, son fiestas religiosas, de la Iglesia, y también está la Feria, que es en abril, que es una semana también, son 7 días. Y en la feria las mujeres nos vestimos con trajes típicos, que se llaman trajes de flamenca, de colores, nos ponemos flores en el pelo y nada, y todo el mundo va a la feria, cantamos. Lo típico es bailar sevillanas, es un baile típico de Sevilla. Y allí la gente se reúne, baila, come, bebe la bebida típica, que es el rebujito, come tortilla de patatas, pimientos fritos, jamón.

María: Y, sobre la Semana Santa, ¿puedes hablar un poco más? ¿Qué es una procesión? ¿Qué hay? ¿Qué podemos ver allí?

Rocío: Pues son — en cada iglesia — son reuniones de personas que sacan a la imagen, al santo, que es o el Cristo o la Virgen, la representación, la sacan por las calles de Sevilla y las personas de esa iglesia los acompañan vestidos de nazareno, acompañan a la imagen por toda Sevilla y la gente sale para verlo, para rezar a la Virgen y pedirle cosas, milagros.

María: ¿Y qué atractivo turístico tiene Sevilla? ¿Qué hace que Sevilla sea atractiva turísticamente? O Andalucía en general.

Rocío: Andalucía. Tiene playas, tiene mucha playa, zona de costa, son muy grandes. Hay muchas discotecas para los jóvenes, que les gusta mucho la zona de la playa. También hace muy buen clima, tenemos también montañas, co que nieva y se puede esquiar y también hacen mucho deporte, surf, mucho deporte en la playa.

¹³⁴⁶ Vgl. Kohn: *Backbone*. Op. cit. (15.02.2018).

31. Arbeitsblatt 22: *Comprensión de texto: Trabajo infantil* – Korpus: Hemero

Comprensión de texto: Trabajo infantil

Lee el texto siguiente y responde a las preguntas debajo.

'Save The Children' ve poco realista tratar de erradicar del trabajo infantil¹³⁴⁷

La ONG defensora de los Derechos de la Infancia 'Save The Children' considera que no es realista promover la erradicación del trabajo infantil y aboga por mejorar las condiciones de los niños que trabajan y van a la escuela.

"No se puede ser radical con la propuesta de erradicar el trabajo infantil porque **primero interesa la supervivencia**, antes que cualquier otra cosa", sostuvo Liliana Orjuela, coordinadora de los programas de promoción del buen trato y prevención de la violencia contra la infancia de 'Save the Children', en un encuentro realizado este miércoles con Europa Press. En el coloquio, convocado con motivo del Día Mundial contra el Trabajo Infantil que se celebra el 12 de junio, participaron expertos y niños que trabajan.

Orjuela advirtió de que una prohibición generalizada del trabajo infantil lleva a perseguir a los que pueden darle la opción a un niño y a una familia de sobrevivir y al propio niño, para quien el trabajo podría suponer "una oportunidad para asumir su vida y su autonomía y sentirse bien apoyando a su familia".

Explotación infantil

La experta **definió el trabajo infantil como "las actividades que realizan los niños y las niñas en beneficio propio, de sus familias y de su comunidad"**, siendo fundamental que las condiciones de trabajo respeten derechos básicos, como la dignidad, la salud y la educación, añadió.

El trabajo en unas condiciones que vulneran esos derechos es explotación y suele constituir situaciones de delito, precisó. **La esclavitud y prácticas similares, como la explotación sexual**, la utilización de los niños para el tráfico de drogas o el reclutamiento forzoso de niños para conflictos armados, **son casos de explotación**.

En referencia a los **conflictos armados**, Victor Bryan Bedriñana Solís, un niño de Perú de 16 años, que trabajó desde pequeño vendiendo pan por las calles de Lima y que actualmente es delegado nacional del Movimiento Nacional de Niños y Adolescentes Trabajadores Organizados de Perú (MNNATSOP), alertó de que la guerrilla de Sendero Luminoso ha intensificado los reclutamientos forzosos.

Orjuela destacó que **"el trabajo infantil no existe sólo en África y Latinoamérica**, sino que existe entre los ricos y los pobres de todos los países". "En España hay niños que trabajan en los mercadillos y que roban", señaló. En el mismo sentido, Romina González, técnica del departamento de protección y promoción de los Derechos de la Infancia de 'Save the Children', recordó que **en España hay niños chinos que ayudan en las tiendas y que no van al colegio**.

¹³⁴⁷ El mundo (2009): op. cit. (17.01.2018).

218 millones de niños trabajan

Según la Organización Internacional del Trabajo (OIT), hay 218 millones de niños trabajadores, de los cuales 160 millones trabajan en las peores formas de trabajo infantil. No obstante, Fabrizio Terenzio, que proporciona apoyo al Movimiento Africano de Niños, Niñas y Adolescentes Trabajadores, subrayó que la OIT no tiene en cuenta que la mayoría de esos niños trabajadores van a la escuela (72 millones no lo hacen).

"Por qué no se habla más que de erradicar de mejorar la condición de quien trabaja y va a la escuela, que es el problema real", se preguntó Terenzio, y criticó que "ideológicamente estamos en una visión en la que **escuela y trabajo** son incompatibles y **hay estadísticas que prueban que son compatibles**".

Asimismo, criticó que la OIT haya incluido trabajos en la lista de las peores formas de trabajo infantil que llevan haciendo los niños desde hace años en muchas comunidades, como ir al pozo a por agua. Orjuela consideró que no hay que catalogar este tipo de trabajos por sectores, sino en función de las condiciones.

Por otro lado, Luis Santos Beltrán Merino, que trabaja con los niños y adolescentes de Perú, alertó de que, "**ante la crisis, el Estado está fomentando más aún el trabajo informal**". Por su parte, Terenzio dijo que los problemas más importantes que está generando la crisis en Senegal son la reducción del turismo, que es la segunda fuente de ingresos del país, y de las remesas, así como la ayuda internacional.

El mundo / Europa Press: Madrid

1. ¿Verdadero (V) o falso (F)?

Marca con una cruz (X) la respuesta correcta y corrige las frases falsas.

	V	F
a) 'Save The Children' es una ONG que defiende los derechos de los niños.		
b) Victor Bryan Bedriñana Solís vendió pan por las calles de Lima durante 16 años.		
c) Victor Bryan Bedriñana Solís es delegado nacional del MNNATSOP.		
d) La guerrilla de Sendero Luminoso obliga a los niños de luchar en conflictos armados.		
e) Estadísticas indican que trabajo y escuela son incompatibles.		
f) El trabajo infantil solo existe en África y Latinoamérica.		

2. Marca con una cruz (X) la respuesta o las respuestas correctas:

- | | |
|---|--|
| a) Con un trabajo digno, el niño puede... | b) Explotación se llama un trabajo que vulnera... |
| <input type="checkbox"/> ... tener más tiempo libre. | <input type="checkbox"/> ... la dignidad. |
| <input type="checkbox"/> ... garantizar su autonomía. | <input type="checkbox"/> ... la educación. |
| <input type="checkbox"/> ... apoyar a su familia. | <input type="checkbox"/> ... la religión. |
| <input type="checkbox"/> ... ir a la escuela. | <input type="checkbox"/> ... la salud. |
| | |
| c) ¿Qué casos de explotación se mencionan? | d) En Senegal la crisis lleva a una reducción de ... |
| <input type="checkbox"/> el tráfico de drogas | <input type="checkbox"/> ... envíos de dinero. |
| <input type="checkbox"/> la prostitución | <input type="checkbox"/> ... ayuda nacional. |
| <input type="checkbox"/> robos | <input type="checkbox"/> ... turismo. |
| <input type="checkbox"/> conflictos armados | <input type="checkbox"/> ... trabajo infantil. |

3. Contesta a las preguntas siguientes en frases completas:

1. ¿Cuándo se celebra el Día Mundial contra el Trabajo Infantil?

2. ¿Quién participó en el coloquio del Día Mundial contra el Trabajo Infantil?

3. ¿Qué es el 'trabajo infantil' exactamente? – Explícalo en tus propias palabras.

4. ¿Cuántos niños trabajadores hay y cuántos de ellos van a la escuela?

4. Explica en tus propias palabras lo que es el mensaje de 'Save The Children' según este artículo.

Comprensión de texto: Trabajo infantil

Lee el texto siguiente y responde a las preguntas debajo.

'Save The Children' ve poco realista tratar de erradicar del trabajo infantil¹³⁴⁸

La ONG defensora de los Derechos de la Infancia 'Save The Children' considera que no es realista promover la erradicación del trabajo infantil y aboga por mejorar las condiciones de los niños que trabajan y van a la escuela.

"No se puede ser radical con la propuesta de erradicar el trabajo infantil porque **primero interesa la supervivencia**, antes que cualquier otra cosa", sostuvo Liliana Orjuela, coordinadora de los programas de promoción del buen trato y prevención de la violencia contra la infancia de 'Save the Children', en un encuentro realizado este miércoles con Europa Press. En el coloquio, convocado con motivo del Día Mundial contra el Trabajo Infantil que se celebra el 12 de junio, participaron expertos y niños que trabajan.

Orjuela advirtió de que una prohibición generalizada del trabajo infantil lleva a perseguir a los que pueden darle la opción a un niño y a una familia de sobrevivir y al propio niño, para quien el trabajo podría suponer "una oportunidad para asumir su vida y su autonomía y sentirse bien apoyando a su familia".

Explotación infantil

La experta **definió el trabajo infantil como "las actividades que realizan los niños y las niñas en beneficio propio, de sus familias y de su comunidad"**, siendo fundamental que las condiciones de trabajo respeten derechos básicos, como la dignidad, la salud y la educación, añadió.

El trabajo en unas condiciones que vulneran esos derechos es explotación y suele constituir situaciones de delito, precisó. **La esclavitud y prácticas similares, como la explotación sexual**, la utilización de los niños para el tráfico de drogas o el reclutamiento forzoso de niños para conflictos armados, **son casos de explotación**.

En referencia a los **conflictos armados**, Victor Bryan Bedriñana Solís, un niño de Perú de 16 años, que trabajó desde pequeño vendiendo pan por las calles de Lima y que actualmente es delegado nacional del Movimiento Nacional de Niños y Adolescentes Trabajadores Organizados de Perú (MNNATSOP), alertó de que la guerrilla de Sendero Luminoso ha intensificado los reclutamientos forzosos.

Orjuela destacó que **"el trabajo infantil no existe sólo en África y Latinoamérica**, sino que existe entre los ricos y los pobres de todos los países". "En España hay niños que trabajan en los mercadillos y que roban", señaló. En el mismo sentido, Romina González, técnica del departamento de protección y promoción de los Derechos de la Infancia de 'Save the Children', recordó que **en España hay niños chinos que ayudan en las tiendas y que no van al colegio**.

¹³⁴⁸ El mundo (2009): op. cit. (17.01.2018).

218 millones de niños trabajan

Según la Organización Internacional del Trabajo (OIT), hay 218 millones de niños trabajadores, de los cuales 160 millones trabajan en las peores formas de trabajo infantil. No obstante, Fabrizio Terenzio, que proporciona apoyo al Movimiento Africano de Niños, Niñas y Adolescentes Trabajadores, subrayó que la OIT no tiene en cuenta que la mayoría de esos niños trabajadores van a la escuela (72 millones no lo hacen).

"Por qué no se habla más que de erradicar de mejorar la condición de quien trabaja y va a la escuela, que es el problema real", se preguntó Terenzio, y criticó que "ideológicamente estamos en una visión en la que **escuela y trabajo** son incompatibles y **hay estadísticas que prueban que son compatibles**".

Asimismo, criticó que la OIT haya incluido trabajos en la lista de las peores formas de trabajo infantil que llevan haciendo los niños desde hace años en muchas comunidades, como ir al pozo a por agua. Orjuela consideró que no hay que catalogar este tipo de trabajos por sectores, sino en función de las condiciones.

Por otro lado, Luis Santos Beltrán Merino, que trabaja con los niños y adolescentes de Perú, alertó de que, **"ante la crisis, el Estado está fomentando más aún el trabajo informal"**. Por su parte, Terenzio dijo que los problemas más importantes que está generando la crisis en Senegal son la reducción del turismo, que es la segunda fuente de ingresos del país, y de las remesas, así como la ayuda internacional.

El mundo / Europa Press: Madrid

1. ¿Verdadero (V) o falso (F)?

Marca con una cruz (X) la respuesta correcta y corrige las frases falsas.	V	F
a) 'Save The Children' es una ONG que defiende los derechos de los niños.	X	
b) Victor Bryan Bedriñana Solís vendió pan por las calles de Lima durante 16 años.		X
c) Victor Bryan Bedriñana Solís es delegado nacional del MNNATSOP.	X	
d) La guerrilla de Sendero Luminoso obliga a los niños de luchar en conflictos armados.	X	
e) Estadísticas indican que trabajo y escuela son incompatibles.		X
f) El trabajo infantil solo existe en África y Latinoamérica.		X

b) No, Victor Bryan Bedriñana Solís tiene 16 años y vendió pan por las calles de Lima desde pequeño.

e) No, las estadísticas prueban que trabajo y escuela son compatibles.

f) No, el trabajo infantil no existe solo en África y Latinoamérica sino en general entre los ricos y los pobres de todos los países (según Orjuela).

2. Marca con una cruz (X) la respuesta o las respuestas correctas:

- | | |
|--|---|
| a) Con un trabajo digno, el niño puede... | b) Explotación se llama un trabajo que vulnera... |
| <input type="checkbox"/> ... tener más tiempo libre. | <input checked="" type="checkbox"/> ... la dignidad. |
| <input checked="" type="checkbox"/> ... garantizar su autonomía. | <input checked="" type="checkbox"/> ... la educación. |
| <input checked="" type="checkbox"/> ... apoyar a su familia. | <input type="checkbox"/> ... la religión. |
| <input type="checkbox"/> ... ir a la escuela. | <input checked="" type="checkbox"/> ... la salud. |
| | |
| c) ¿Qué casos de explotación se mencionan? | d) En Senegal la crisis lleva a una reducción de ... |
| <input checked="" type="checkbox"/> el tráfico de drogas | <input checked="" type="checkbox"/> ... envíos de dinero. |
| <input checked="" type="checkbox"/> la prostitución | <input type="checkbox"/> ... ayuda nacional. |
| <input type="checkbox"/> robos | <input checked="" type="checkbox"/> ... turismo. |
| <input checked="" type="checkbox"/> conflictos armados | <input type="checkbox"/> ... trabajo infantil. |

3. Contesta a las preguntas siguientes en frases completas:

1. ¿Cuándo se celebra el Día Mundial contra el Trabajo Infantil?
Se celebra el 12 de junio.
2. ¿Quién participó en el coloquio del Día Mundial contra el Trabajo Infantil?
Participaron expertos y niños que trabajan.
3. ¿Qué es el 'trabajo infantil' exactamente? – Explícalo en tus propias palabras.
Son las actividades que hacen los niños para ellos mismos, sus familias y su comunidad.
4. ¿Cuántos niños trabajadores hay y cuántos de ellos van a la escuela?
Hay 218 millones niños trabajadores y de estos 146 van a la escuela.

4. Explica en tus propias palabras lo que es el mensaje de 'Save The Children' según este artículo.

'Save the Children' dice que no es posible erradicar el trabajo infantil. Entonces intentan mejorar la condición de los niños que trabajan y van a la escuela.

32. Arbeitsblatt 23: *El Español de Argentina* – Korpus CORPES**El Español de Argentina****1. El Voseo: Pronombres y sus formas verbales**

Encuentra la forma correcta para la segunda persona singular en el español de Argentina y rellena la tabla siguiente.

Zunächst öffnest du die Internetseite des *Corpus del Español del Siglo XXI* – CORPES (<http://web.frl.es/CORPES/view/inicioExterno.view>) und gibst das Verb (z. B.: *vivir*) als *lema* ein. Daraufhin wählst du *verbo* im Bereich *clase de palabra* als Wortart sowie *indicativo* und *presente* als Modus und Tempus aus. Nach einem Klick auf das Feld *Proximidad* eröffnet sich eine zweite Eingabezeile. Hier gibst du *vos* als *forma* ein, wählst *pronombre personal* als Wortart sowie die zweite Person Singular als Person und Numerus aus. Am besten suchst du in einem Intervall von einem Wort auf der linken Seite.

The screenshot shows the CORPES search interface. The search criteria are: Lema: vivir, Forma: vos, Clase de palabra: pronombre personal, Modo: indicativo, Tiempo: presente, Persona: segunda, Número: singular. The search results show 12 cases in 11 documents. The results are listed in a table with columns: REF. (Clasificación, país), CONCORDANCIA, and Ordenar por: Año ascendente.

REF. (Clasificación, país)	CONCORDANCIA	Ordenar por: Año ascendente
1 2002 Par.	hablando sentados a la mesa de la cocina, y tu madre: "¡Dejen el pan!" Joaquín, "vos vivís en mi casa y yo soy el más pobre del colegio.	
2 2003 Arg.	pedos, te torturan -con un velador de tu casa te hacen una picana eléctrica-. Vos vivís ahí adentro porque sos gerente de un banco,	
3 2003 Arg.	CHICO: Vos que sabés...vos vivís allá adentro.	
4 2003 Arg.	de infantes, con todos esos niños de ojitos crueles, preguntándose: "¿Por qué vos vivís con tus tíos? ¿Y tu papá y tu mamá dónde est	
5 2004 Guat.	un autobús extraurbano, en Chichicastenango, subió un anciano que le dijo: "Vos vivís en una casa con corredor grande y a la par de	
6 2006 Col.	pregunta si es hora de sacar las empanadas del horno. Sí, hay un horno en tu casa y vos vivís en esta casa aunque no parezca, y él sólo díc	
7 2006 Arg.	vos estás, de la gente con la que convivís, porque es toda una rutina lo que vos vivís todos los días. Te levantás, ves a la misma ge	
8 2009 Hond.	cadáveres. Bruta que sos. Jetabierta. Los embajadores viven más allá de la luna y vos vivís en el lodo tiundo en que vivís, mi princesita de	

CORPES: Suche nach der Verbform der 2. Person Singular von *vivir* im *voseo*.

Verbo	España	Argentina	Verbo	España	Argentina
vivir	tú vives	vos vivís	andar	tú _____	vos _____
comer	tú _____	vos _____	ir	tú _____	vos _____
trabajar	tú _____	vos _____	conocer	tú _____	vos _____
cantar	tú _____	vos _____	decir	tú _____	vos _____
leer	tú _____	vos _____	poder	tú _____	vos _____
ser	tú _____	vos _____	salir	tú _____	vos _____
estar	tú _____	vos _____	dormir	tú _____	vos _____
querer	tú _____	vos _____	coger	tú _____	vos _____
ver	tú _____	vos _____	pensar	tú _____	vos _____
saber	tú _____	vos _____	sentir	tú _____	vos _____
tener	tú _____	vos _____	venir	tú _____	vos _____

2. Diferencias en la pronunciación

Elige la opción correcta para la pronunciación de las palabras siguientes.

Verwende auch hier das CORPES (<http://www.rae.es/recursos/banco-de-datos/corpes-xxi>), beschränke die Suche bei dieser Aufgabe allerdings auf das Subkorpus *Oral*, indem du zunächst auf das Feld *Subcorpus* klickst und dort im Bereich *Medio* die Option *Oral* auswählst. Daraufhin gibst du eines der untenstehenden Wörter (z. B.: *mayo*) als *Lema* ein, klickst auf das Feld *Concordancia* und wählst ein Beispiel aus Spanien (*Esp.*) oder Argentinien (*Arg.*) aus.

Concordancias | Coapariciones | Configuración | Ayuda | Modo de cita | Sugerencias

Lema: Forma: Clase de palabra: ☐ Grafía original

Subcorpus:

Título: Autor: Fecha de clasificación:

Origen:
 (Todos)
 América
 España
 Filipinas

Medio:
 Escrito
 Oral

Tipología:
 Debate
 Discurso
 Entrevista

Sexo:
 Hombre
 Mujer
 No indicado

Grupo de edad:
 0-14
 15-19
 20-34

Marcas: ☐ Cursiva ☐ Negrita ☐ Subrayado ☐ Versalitas

68 casos en 47 documentos.

REF.	(Clasificación, país)	CONCORDANCIA	Ordenar por:	Año ascendente
1	2001 Esp.	en / en / a finales de mayo por ejemplo		
2	2001 Esp.	muchos los alicientes que brinda la capital al visitante a lo largo de todo el mes de mayo		
3	2001 Esp.	tumoral // pero es más / hoy se ha comentado aquí que mmm / ahora / en el próximo mes de mayo se van a presentar en el // en el Congreso		
4	2001 Esp.	artículos siguen una regla determinada // el día / en este caso el doce / se une al mes / mayo / mediante la preposición de / y decimos //		
5	2001 Cuba	universitario no les ha // eeh posibilitado esto // y después // el segundo nivel / marzo abril y mayo / la especialización // esos cursos tienen /		
6	2001 Esp.	/ y muchas de las medidas se pondrán en marcha o ya se han puesto en marcha entre mayo de este año y diciembre de este mismo año		
7	2001	los aceites de orujo // asimismo / tuvieron ustedes conocimiento en los últimos días de mayo o primeros de junio de que la República C		
8	2001 Esp.	yo espero que se pueda aprobar en una próxima reunión / yo creo que es a fines de mayo la próxima convocatoria para los ministros		

CORPES: Suche nach dem Wort *mayo* im mündlichen Subkorpus.

Nach einem Klick auf das jeweilige Beispiel wird unter den Konkordanzen der dazugehörige Textausschnitt inklusive einer Audiodatei angezeigt. Hör dir ein spanisches sowie ein argentinisches Beispiel an, und wähle die jeweils korrekte Aussprache in der Tabelle aus.

Palabra	España	Argentina	Palabra	España	Argentina
mayo	[j]	[ɣ]	aparecer	[θ]	[θ]
	[h]	[h]		[s]	[s]
	[ɜ] / [f]	[ɜ] / [ɟ]		[z]	[z]
ayer	[j]	[j]	inteligencia	[θ]	[θ]
	[h]	[h]		[s]	[s]
	[ɜ] / [ɟ]	[ɜ] / [ɟ]		[z]	[z]
llama	[j]	[j]	zapatos	[θ]	[θ]
	[lj]	[lj]		[s]	[s]
	[ɜ] / [ɟ]	[ɜ] / [ɟ]		[z]	[z]

El Español de Argentina – Solución

1. El Voseo: Pronombres y sus formas verbales

Encuentra la forma correcta para la segunda persona singular en el español de Argentina y rellena la tabla siguiente.

Zunächst öffnest du die Internetseite des *Corpus del Español del Siglo XXI – CORPES* (<http://web.frl.es/CORPES/view/inicioExterno.view>) und gibst das Verb (z. B.: *vivir*) als *lema* ein. Daraufhin wählst du *verbo* im Bereich *clase de palabra* als Wortart sowie *indicativo* und *presente* als Modus und Tempus aus. Nach einem Klick auf das Feld *Proximidad* eröffnet sich eine zweite Eingabezeile. Hier gibst du *vos* als *forma* ein, wählst *pronombre personal* als Wortart sowie die zweite Person Singular als Person und Numerus aus. Am besten suchst du in einem Intervall von einem Wort auf der linken Seite.

The screenshot shows the CORPES search interface. The top navigation bar includes links for ConcordanCIAS, Coapariciones, Configuración, Ayuda, Modo de cita, and Sugerencias. The search form is divided into two sections. The first section has fields for Lema (vivir), Forma (empty), Clase de palabra (verbo), and Modos (indicativo, presente). The second section has fields for Lema (empty), Forma (vos), Clase de palabra (pronombre personal), and Modos (Distancia, Intervalo). The search results section shows 12 cases in 11 documents, with a table of results including the reference, classification, and the concordance.

REF.	(Clasificación, país)	CONCORDANCIA	Ordenar por:	Año ascendente
1	2002 Par.	hablando sentados a la mesa de la cocina, y tu madre: "¡Dejen el pan!" Joaquín, "vos vivís en mi casa y yo soy el más pobre del colegio.		
2	2003 Arg.	pedos, te torturan -con un velador de tu casa te hacen una picana eléctrica-. Vos vivís ahí adentro porque sos gerente de un banco,		
3	2003 Arg.	CHICO: Vos que sabés...vos vivís allá adentro.		
4	2003 Arg.	de infantes, con todos esos niños de ojitos crueles, preguntándome: "¿Por qué vos vivís con tus tíos? ¿Y tu papá y tu mamá dónde est		
5	2004 Guat.	un autobús extraurbano, en Chichicastenango, subió un anciano que le dijo: "Vos vivís en una casa con corredor grande y a la par de		
6	2006 Col.	pregunta si es hora de sacar las empanadas del horno. Sí, hay un horno en tu casa y vos vivís en esta casa aunque no parezca, y él sólo dijo		
7	2006 Arg.	vos estás, de la gente con la que convivís, porque es toda una rutina lo que vos vivís todos los días. Te levantás, ves a la misma ge		
8	2009 Hond.	cadáveres. Bruta que sos. Jetabierta. Los embajadores viven más allá de la luna y vos vivís en el lodo tiundo en que vivís, mi princesita de		

CORPES: Suche nach der Verbform der 2. Person Singular von *vivir* im *voseo*.

Verbo	España	Argentina	Verbo	España	Argentina
vivir	tú vives	vos vivís	andar	tú andas	vos andás
comer	tú comes	vos comés	ir	tú vas	vos vas
trabajar	tú trabajas	vos trabajás	conocer	tú conoces	vos conocés
cantar	tú cantas	vos cantás	decir	tú dices	vos decís
leer	tú lees	vos leés	poder	tú puedes	vos podés
ser	tú eres	vos sos	salir	tú sales	vos salís
estar	tú estás	vos estás	dormir	tú duermes	vos dormís
querer	tú quieres	vos querés	coger	tú coges	vos cogés
ver	tú ves	vos ves	pensar	tú piensas	vos pensás
saber	tú sabes	vos sabés	sentir	tú sientes	vos sentís
tener	tú tienes	vos tenés	venir	tú vienes	vos venís

2. Diferencias en la pronunciación

Elige la opción correcta para la pronunciación de las palabras siguientes.

Verwende auch hier das CORPES (<http://www.rae.es/recursos/banco-de-datos/corpes-xxi>), beschränke die Suche bei dieser Aufgabe allerdings auf das Subkorpus *Oral*, indem du zunächst auf das Feld *Subcorpus* klickst und dort im Bereich *Medio* die Option *Oral* auswählst. Daraufhin gibst du eines der untenstehenden Wörter (z. B.: *mayo*) als *Lema* ein, klickst auf das Feld *Concordancia* und wählst ein Beispiel aus Spanien (*Esp.*) oder Argentinien (*Arg.*) aus.

Concordancias | Coapariciones | Configuración | Ayuda | Modo de cita | Sugerencias

Lema: Forma: Clase de palabra: ☐ Grafía original

Subcorpus:

Título: Autor: Fecha de clasificación:

Origen:
 (Todos)
 América
 España
 Filipinas

Medio:
 Escrito
 Oral

Tipología:
 Debate
 Discurso
 Entrevista

Sexo:
 Hombre
 Mujer
 No indicado

Grupo de edad:
 0-14
 15-19
 20-34

Marcas: ☐ Cursiva ☐ Negrita ☐ Subrayado ☐ Versalitas

68 casos en 47 documentos.

REF.	(Clasificación, país)	CONCORDANCIA	Ordenar por:	Año ascendente
1	2001 Esp.	en / en / a finales de mayo por ejemplo		
2	2001 Esp.	muchos los alicientes que brinda la capital al visitante a lo largo de todo el mes de mayo		
3	2001 Esp.	tumoral // pero es más / hoy se ha comentado aquí que mmm / ahora / en el próximo mes de mayo se van a presentar en el // en el Congreso		
4	2001 Esp.	artículos siguen una regla determinada // el día / en este caso el doce / se une al mes / mayo / mediante la preposición de / y decimos //		
5	2001 Cuba	universitario no les ha // eeh posibilitado esto // y después // el segundo nivel / marzo abril y mayo / la especialización // esos cursos tienen /		
6	2001 Esp.	/ y muchas de las medidas se pondrán en marcha o ya se han puesto en marcha entre mayo de este año y diciembre de este mismo año		
7	2001	los aceites de orujo // asimismo / tuvieron ustedes conocimiento en los últimos días de mayo o primeros de junio de que la República C		
8	2001 Esp.	yo espero que se pueda aprobar en una próxima reunión / yo creo que es a fines de mayo la próxima convocatoria para los ministros		

CORPES: Suche nach dem Wort *mayo* im mündlichen Subkorpus.

Nach einem Klick auf das jeweilige Beispiel wird unter den Konkordanzen der dazugehörige Textausschnitt inklusive einer Audiodatei angezeigt. Hör dir ein spanisches sowie ein argentinisches Beispiel an, und wähle die jeweils korrekte Aussprache in der Tabelle aus.

Palabra	España	Argentina	Palabra	España	Argentina
mayo	[j]	[j]	aparecer	[θ]	[θ]
	[h]	[h]		[s]	[s]
	[3]/[ʃ]	[3]/[ʃ]		[z]	[z]
ayer	[j]	[j]	inteligencia	[θ]	[θ]
	[h]	[h]		[s]	[s]
	[3]/[ʃ]	[3]/[ʃ]		[z]	[z]
llama	[j]	[j]	zapatos	[θ]	[θ]
	[h]	[h]		[s]	[s]
	[3]/[ʃ]	[3]/[ʃ]		[z]	[z]

33. Arbeitsblatt 24: *Los idiomas de España* – Korpus: CLUVI

Los idiomas de España: Castellano, Gallego, Catalán y Vasco (Euskara)

Verwende das Subkorpus CONSUMER des CLUVI (<http://sli.uvigo.es/CLUVI/index.php?corpus=20&tipo=26&lang=en>), um dieses Arbeitsblatt auszufüllen. Zunächst gibst du einen Suchbegriff in diejenige Zeile ein, deren Sprache er entstammt (z. B.: *libro* in die spanische Eingabezeile), und klickst auf *Search*. In den daraufhin erscheinenden Ergebnissen suchst du die Entsprechungen des Suchbegriffs in den anderen Sprachen.

Select texts ▾

CONSUMER ES-GL-CA-EU

Spanish
libro

Galician
Galician text

Catalan
Catalan text

Basque
Basque text

Search →

☐ Show marks

CLUVI: Eingabemaske des Subkorpus CONSUMER.

1. Busca seis palabras españolas en el corpus CONSUMER y escribe los equivalentes gallegos, catalanes y vascos en las columnas correspondientes:

ES El tesoro puede tomar formas muy diversas, pero siempre contará con un **libro** de visitas.

GL O tesouro pode tomar formas moi diversas, pero sempre contará cun **libro** de visitas.

CA El tresor pot prendre formes molt diverses, però sempre disposarà d'un **llibre** de visites.

EU Altxorra era askotakoa izan daiteke baina bisitarien **liburua** izango du beti.

CLUVI: Ergebnis der Suche nach dem spanischen Substantiv *libro* im Subkorpus CONSUMER.

español	gallego	catalán	vasco (euskara)

a) ¿Cuáles de estos idiomas se parecen?

b) ¿Qué idioma es totalmente diferente?

- c) ¿Qué es el problema cuando quieres escribir los equivalentes del idioma buscado en ejercicio b) en la columna?

2. ¿Qué texto está escrito en qué idioma?

_____	_____	_____	_____
També els equips de futbol organitzen campus i escoles d'estiu per als més petits.	Tamén os equipos de fútbol organizan campus e escolas de verán para os máis pequenos.	Etxeko txikientzako udako kanpus eta eskolak futbol taldeek ere antolatzen dituzte.	También los equipos de fútbol organizan campus y escuelas de verano para los más pequeños.

3. ¿De qué idioma se trata en las frases siguientes? ¿Qué significa en español?

- a) Naturarengana gerturatu eta beraz onura ateratzeko hoberena bizikletazko edo oinezko txangoak dira. ➔ Idioma: _____

Español:

- b) El nostre temps lliure el dediquem preferentment a visitar els centres comercials o a anar d'aparador en aparador. ➔ Idioma: _____

Español:

- c) A escola non só transmite coñecementos, tamén ten a función de formar persoas complementando a educación que reciben dos pais. ➔ Idioma: _____

Español:

Los idiomas de España: Castellano, Gallego, Catalán y Vasco (Euskara) – *Solución*

Verwende das Subkorpus CONSUMER des CLUVI (<http://sli.uvigo.es/CLUVI/index.php?corpus=20&tipo=26&lang=en>), um dieses Arbeitsblatt auszufüllen. Zunächst gibst du einen Suchbegriff in diejenige Zeile ein, deren Sprache er entstammt (z. B.: *libro* in die spanische Eingabezeile), und klickst auf *Search*. In den daraufhin erscheinenden Ergebnissen suchst du die Entsprechungen des Suchbegriffs in den anderen Sprachen.

1. Busca seis palabras españolas en el corpus CONSUMER y escribe los equivalentes gallegos, catalanes y vascos en las columnas correspondientes:

Select texts ▾

CONSUMER ES-GL-CA-EU

Spanish
libro

Galician
Galician text

Catalan
Catalan text

Basque
Basque text

Search →

☐ Show marks

CLUVI: Eingabemaske des Subkorpus CONSUMER.

ES El tesoro puede tomar formas muy diversas, pero siempre contará con un **libro** de visitas.

GL O tesouro pode tomar formas moi diversas, pero sempre contará cun **libro** de visitas.

CA El tresor pot prendre formes molt diverses, però sempre disposarà d'un **llibre** de visites.

EU Altxorra era askotakoa izan daiteke baina bisitarien **liburua** izango du beti.

CLUVI: Ergebnis der Suche nach dem spanischen Substantiv *libro* im Subkorpus CONSUMER.

español	gallego	catalán	vasco (euskara)
libro	libro	llibre	<i>liburua, liburu, liburuak, ...</i>
día	día	dia	<i>egun, eguneko, egunero, ...</i>
perro	can	gos	<i>txakurra txakurrak, txakurrari, ...</i>
estudiar	estudiar, estudar	estudiar	<i>aztertu, aztertze, aztertzea, ...</i>
interesante	interesante	interessant	<i>interesgarri, interesgarria, ...</i>
videojuegos	videoxogos	videojocs	<i>bideojokoen, bideojokoak, ...</i>

- a) ¿Cuáles de estos idiomas se parecen?

El español y el gallego se parecen.

- b) ¿Qué idioma es totalmente diferente?

El vasco es totalmente diferente.

- c) ¿Qué es el problema cuando quieres escribir los equivalentes del idioma buscado en ejercicio b) en la columna?

El vasco tiene una estructura totalmente diferente. Encontrar el equivalente vasco de una palabra castellana es muy difícil porque las palabras vascas cambian la terminación en contextos diferentes, por ejemplo, cuando se usa otra preposición en castellano.

2. ¿Qué texto está escrito en qué idioma?

<u>catalán</u>	<u>gallego</u>	<u>euskara (vasco)</u>	<u>español</u>
També els equips de futbol organitzen campus i escoles d'estiu per als més petits.	Tamén os equipos de fútbol organizan campus e escolas de verán para os máis pequenos.	Etxeko txikientzako udako kanpus eta eskolak futbol taldeek ere antolatzen dituzte.	También los equipos de fútbol organizan campus y escuelas de verano para los más pequeños.

3. ¿De qué idioma se trata en las frases siguientes? ¿Qué significa en español?

- a) Naturarengana gerturatu eta beraz onura ateratzeko hoberena bizikletazko edo oinezko txangoak dira. → Idioma: vasco

Español: Las marchas a pie o en bicicleta son el mejor medio para aproximarse y disfrutar de la naturaleza.

- b) El nostre temps lliure el dediquem preferentment a visitar els centres comercials o a anar d'aparador en aparador. → Idioma: catalán

Español: Nuestro tiempo libre lo dedicamos preferentemente a visitar los centros comerciales o ir de escaparate en escaparate.

- c) A escola non só transmite coñecementos, tamén ten a función de formar persoas complementando a educación que reciben dos pais. → Idioma: gallego

Español: La escuela no sólo trasmite conocimientos, también tiene la función de formar personas complementando la educación que reciben de los padres.

34. Arbeitsblatt 25: *Repaso: Estar + gerundio* – Korpus: *Adelante-Korpus*

Repaso: Estar + gerundio

Pon los verbos entre paréntesis en la forma correcta.

1. Elena: Ah, aquí llega Lucía. Hola, Lucía, ¿qué tal?
 Lucía: Bien, bien. _____ (buscar) a Javier. ¿Sabes dónde está?
 Elena: Está en la recepción con Juan. _____ (atender) a clientes.
 Tim: ¿Y quién es Javier?
 Lucía: Es nuestro jefe de recepción. Se llama Javier Suárez. Mira, es este señor alto, delgado y moreno. Ahora _____ (escribir) algo en un papel. Trabaja en este hotel desde hace muchos años...
2. Chico: ¿Y Javier dónde está?
 Chica: ¿Javier? No sé. ¿No _____ (atender) a unos clientes en la recepción?
 Chico: No, en la recepción no está.
3. Es que estuvimos en un concierto de rock y ahora ya no hay autobús. _____ (buscar) un hotel.
4. Locutor: Hoy en nuestro programa de radio _____ (hablar) con jóvenes europeos que _____ (estudiar) o _____ (trabajar) aquí en Madrid. Vamos a empezar hablando con Susanne.
5. Lena tiene diecinueve años, es alemana y _____ (hacer) unas prácticas en León. Ella habla muy bien español, francés e inglés.
6. Otra cosa, si ves que lo que haces no _____ (salir) bien, no tengas miedo de improvisar y haz otra cosa.
7. Madre: ¡Roberto! ¿Dónde estás?
 Chico: En la cocina.
 Madre: ¿_____ (preparar) el desayuno?
 Chico: Noooo. _____ (lavar) los platos.
 Madre: ¡Ah! Muy bien hijo, muy bien.
 Chico: ¡Madres!
8. Bueno, ahora voy a descansar. ¿Mañana me contarás de esa ONG que te _____ (ayudar)?
9. Hola, Luis: No tengo mucho tiempo para escribir, pero quería contarte que lo _____ (pasar) muy bien en México. Nuestro hotel es genial, la comida es muy auténtica.

10. Hay una manifestación. Parece que quieren construir un hotel en la zona y la gente _____ (protestar). Claro, imagina, quieren que el hotel tenga habitaciones para mil personas.
11. Acabo de leer que en la universidad Miguel Hernández de Elche hay investigadores que _____ (crear) un césped artificial para los campos de golf.
12. La llegada de gran número de turistas a nuestra región _____ (causar) dolores de cabeza a nuestros servicios policiales.
13. Y en cuanto vuelvo de un viaje, mis jefes ya _____ (preparar) otro. Por eso no tengo mucho tiempo para mis amigos o mi familia.
14. Julio: ¿Cuándo? Yo también tuve que ir solo. Te _____ (esperar) y nunca llegaste.
Sergio: ¿El viernes?
Julio: No, el sábado.
15. _____ (escribir) un libro con recetas, porque también le gustaba escribir.
16. Puede hacer turismo en las preciosas haciendas cafeteras de la turística Ruta del Café. Colombia _____ (ofrecer) ahora con su producto más típico, el café, la mejor cara del país.

Repaso: Estar + gerundio – Solución

Pon los verbos entre paréntesis en la forma correcta.

1. Ah, aquí llega Lucía. Hola, Lucía, ¿qué tal?
Lucía: Bien, bien. Estoy buscando (buscar) a Javier. ¿Sabes dónde está?
Elena: Está en la recepción con Juan. Están atendiendo (atender) a clientes.
Tim: ¿Y quién es Javier?
Lucía: Es nuestro jefe de recepción. Se llama Javier Suárez. Mira, es este señor alto, delgado y moreno. Ahora está escribiendo (escribir) algo en un papel. Trabaja en este hotel desde hace muchos años...
2. Chico: ¿Y Javier dónde está?
Chica: ¿Javier? No sé. ¿No está atendiendo (atender) a unos clientes en la recepción?
Chico: No, en la recepción no está.
3. Es que estuvimos en un concierto de rock y ahora ya no hay autobús. Estamos buscando (buscar) un hotel.
4. Locutor: Hoy en nuestro programa de radio estamos hablando (hablar) con jóvenes europeos que están estudiando (estudiar) o trabajando (trabajar) aquí en Madrid. Vamos a empezar hablando con Susanne.
5. Lena tiene diecinueve años, es alemana y está haciendo (hacer) unas prácticas en León. Ella habla muy bien español, francés e inglés.
6. Otra cosa, si ves que lo que haces no está saliendo (salir) bien, no tengas miedo de improvisar y haz otra cosa.
7. Madre: ¡Roberto! ¿Dónde estás?
Chico: En la cocina.
Madre: ¿Estás preparando (preparar) el desayuno?
Chico: Noooo. Estoy lavando (lavar) los platos.
Madre: ¡Ah! Muy bien hijo, muy bien.
Chico: ¡Madres!
8. Bueno, ahora voy a descansar. ¿Mañana me contarás de esa ONG que te está ayudando (ayudar)?
9. Hola, Luis: No tengo mucho tiempo para escribir, pero quería contarte que lo estamos pasando (pasar) muy bien en México. Nuestro hotel es genial, la comida es muy auténtica.
10. Hay una manifestación. Parece que quieren construir un hotel en la zona y la gente está protestando (protestar). Claro, imagina, quieren que el hotel tenga habitaciones para mil personas.
11. Acabo de leer que en la universidad Miguel Hernández de Elche hay investigadores que están creando (crear) un césped artificial para los campos de golf.

12. La llegada de gran número de turistas a nuestra región está causando (causar) dolores de cabeza a nuestros servicios policiales.
13. Y en cuanto vuelvo de un viaje, mis jefes ya están preparando (preparar) otro. Por eso no tengo mucho tiempo para mis amigos o mi familia.
14. Julio: ¿Cuándo? Yo también tuve que ir solo. Te estuve esperando (esperar) y nunca llegaste.
Sergio: ¿El viernes?
Julio: No, el sábado.
15. Estaba escribiendo (escribir) un libro con recetas, porque también le gustaba escribir.
16. Puede hacer turismo en las preciosas haciendas cafeteras de la turística Ruta del Café. Colombia está ofreciendo (ofrecer) ahora con su producto más típico, el café, la mejor cara del país.

35. Arbeitsblatt 26: *Argentina: Vocabulario más frecuente* - Korpus: *Argentina-Korpus***Argentina: Vocabulario más frecuente**

mate	laura	gobierno	hijo	tomar	nacer
argentina	hinchas	justicia	infusión	alba	partir
años	pepe	millón	ir	antiguo	pasar
boca	llamar	nacional	medio	bebida	plata
américa	país	reconocer	niño	café	poder
mayo	militar	latino	puerto	caliente	preguntar
mundo	vivir	plaza	saber	chile	república
democracia	importante	popular	santo	conocer	river-ski
casa	ciudad	presidente	venir	crear	río
ramón	club	selección	yerba	cultural	secuestrar
agua	dictadura	siglo	campeonato	deporte	social
pepón	capital	tango	corriente	gente	té
fútbol	derecho	centro	formar	historia	asociación
abuelas	barrio	convertir	junior	histórico	atracción
madre	calle	dudar	llegar	importancia	auto-
mundial	día	estadio	nombre	internacional	regulación
aires	equipo	estar	padre	jugar	baile
copa	habitante	ganar	paso	mafalda	calabaza
personas	avenir	gustar	salir	museo	clase

36. Arbeitsblatt 27: *Chile: Vocabulario más frecuente* – Korpus: *Chile-Korpus*

Chile: Vocabulario más frecuente

chile	interesar	sueño	peso	vivir	julio
año	fundo	violación	salvador	volver	libertad
país	laboral	entregar	santo	augusto	marzo
estar	presidente	hablar	minar	cambio	minero
gobierno	humano	pedir	condicionar	celular	mixto
allende	realidad	comuna	militar	clima	moderno
chileno	tiempo	derecho	avanzar	compraventa	muerto
contraste	historia	dns	sustituir	conadi	principio
dictadura	mundo	estudiante	ubicar	conflicto	retroceso
mapuche	reformar	pinochet	sur	conocido	septiembre
tierra	curacautín	socialista	constitución	cultural	siglo
económico	economía	sucedido	democrático	derrumbe	sociedad
población	extraño	venir	día	especialmente	trabajo
golpe	hectárea	viaje	dónde	exilio	vendedor
comunidad	indígena	caso	geografía	falso	vuelta
santiago	líder	joven	josé	hermano	zona
millón	pérdida	verdad	músico	hospitalidad	cobre
minería	quizás	octubre	seguir	izquierdo	hacer

37. Arbeitsblatt 28: *Chile: Vocabulario nuevo* – Korpus: *Chile-Korpus***Chile: Vocabulario nuevo**

hubo	carcajadas	décadas	lanco	representar
mapuche	casos	economías	litio	rescatados
minería	causó	emotiva	maestros	resguardar
extraño	caótica	empleadores	managers	retirado
socialista	celebraciones	entenderá	manipulaba	revolucionario
sucedido	celular	entrará	mayoritariamente	revés
violaciones	celulares	entregadas	mediterráneo	rivera
avance	cerrar	entrevistas	mendigos	ríen
comuna	chilena	envuelto	metrópolis	sacaron
derrumbe	civilizado	escasa	modifica	santiaguinos
estaciones	claramente	escolares	modificar	seguramente
familiar	clasicista	escondía	motivacionales	sentía
fundo	cliché	esfumaron	mínimo	significó
fundos	comerse	espejos	músico	sobrepasa
hospitalidad	comió	esquina	nacionalización	sobrepasó
mapuches	comparable	estuvieron	neo	solucionará
mina	compraron	experimentado	neoliberalismo	sucede
mineros	comunidades	extravertido	novedad	sucedieron
mixta	conflicto	falso	objetivos	sufrieron
pedía	conflictos	farándula	obligó	sílabas
resulta	cordillerana	financiados	obstante	taxistas
retroceso	corporación	fiscalización	oficinistas	televisivos
torturados	cosmopolita	garantizar	olla	temperamento
acerca	creció	gravemente	opiniones	temperaturas
acercamiento	criminales	gris	oreja	templadas
actuar	curiosidad	guardó	palacios	terribles
afrontar	curiosos	haremos	papelito	territorial
agraria	cámaras	hecha	patricia	toquen
alarma	daba	hemisferio	pegado	trajes
alrededor	debía	hollywoodenses	perder	transcurrido
ambas	definidas	héroes	periódico	transformándose
ambulante	deje	húmedas	personalmente	trató
apoyaban	demandar	importunan	pidan	tremendo
apoyaran	desaparecieron	impresiones	pierdo	ubicada
apurados	desconocida	impresión	pintando	ubicados
arcadas	descontento	inaceptables	predio	unidad
asesinadas	destacado	independizó	prensa	vagar
asesores	destierro	indicado	privilegios	vecino
autoridades	devuelvan	indicó	produjo	viendo
avanzadas	dictando	influir	provocó	visitante
banderas	dieciocho	intenciones	psicóloga	vista
basta	directora	intervención	ratón	vivieron
bicentenario	discusiones	inversión	realizada	viña
bisutería	discute	invirtieron	refiere	volverá
boicotearon	diseño	invirtió	reinventado	votaciones
cambia	docencia	justas	reivindicación	yuppies

38. Arbeitsblatt 29: *Argentina: Las preposiciones 'por' y 'para' – Korpus: Argentina-Korpus*

Argentina: Las preposiciones *por* y *para*

Completa las frases siguientes con la preposición correcta.

1. Hasta el siglo XX muchos países de América Latina pasaron _____ años de violencia y de dictaduras militares.
2. La educación a distancia será mucho más fácil y barata. ¿Se comunicarán estos niños con sus profesores _____ internet?
3. Siete ciudades son megalópolis. De ellas _____ ejemplo México, Buenos Aires y Caracas están en América Latina.
4. ¿Cómo cambiará la vida en estas megalópolis? ¿Serán un modelo _____ otras ciudades?
5. El Club de fútbol Boca Juniors es el más popular del país y cuenta con fanáticos _____ todo el mundo.
6. El estadio está en el barrio de La Boca en Buenos Aires y tiene espacio _____ 60.000 personas.
7. El tango argentino es un baile de pareja y un género musical que se distingue _____ su carácter nostálgico y apasionado.
8. Mafalda es una niña de clase media argentina con una imagen del mundo muy especial. Se interesa _____ el mundo, el medio ambiente y _____ la política, pero odia la sopa.
9. Desde 1991 se disputa un Campeonato de Fútbol Femenino y en 2003 la selección de fútbol argentino se clasificó _____ primera vez _____ la Copa Mundial.
10. El fútbol fue introducido en Argentina a finales del siglo XIX _____ inmigrantes británicos.
11. Hoy existen dos ligas nacionales: la Primera División y la Nacional B con dos campeonatos _____ año.
12. Dice una leyenda que los fundadores del club decidieron tomar la bandera del primer barco que pasara _____ el río Riachuelo como colores de su club. Fue un barco sueco _____ eso sus colores son azul y amarillo.
13. El Estadio Monumental pertenece a River Plate y allí juega también la selección nacional. Además, es un lugar _____ conciertos internacionales de rock y pop.
14. La yerba mate son hojas secas, ligeramente tostadas y picadas o molidas. _____ la preparación del mate normalmente se llena la mitad de la calabaza con yerba.
15. El mate frío es una bebida típica _____ los meses de calor.

8. Anhang

16. El barrio de Recoleta es famoso _____ su cementerio y su centro cultural.
17. En Argentina y Chile la época de la dictadura se caracterizó _____ la violación sistemática de los derechos humanos: secuestros, arrestos, torturas, y extorsiones.
18. Yo, yo he comprendido hasta qué punto callar es importante. Ahora, _____ las noches, después de comer, mi abuela me lleva siempre a casa de Carlitos, su hermano.
19. Se para, nuevamente. Y nosotros lo abandonamos apenas arrimado a la vereda, y nos perdemos _____ las calles transversales, sin mirar una sola vez atrás.
20. Uno de los mayores problemas del país es la desocupación. Hoy el 10,7 _____ ciento de la población activa está sin trabajo.

Argentina: Las preposiciones *por* y *para* – Solución

Completa las frases siguientes con la preposición correcta.

1. Hasta el siglo XX muchos países de América Latina pasaron por años de violencia y de dictaduras militares.
2. La educación a distancia será mucho más fácil y barata. ¿Se comunicarán estos niños con sus profesores por internet?
3. Siete ciudades son megalópolis. De ellas por ejemplo México, Buenos Aires y Caracas están en América Latina.
4. ¿Cómo cambiará la vida en estas megalópolis? ¿Serán un modelo para otras ciudades?
5. El Club de fútbol Boca Juniors es el más popular del país y cuenta con fanáticos por todo el mundo.
6. El estadio está en el barrio de La Boca en Buenos Aires y tiene espacio para 60.000 personas.
7. El tango argentino es un baile de pareja y un género musical que se distingue por su carácter nostálgico y apasionado.
8. Mafalda es una niña de clase media argentina con una imagen del mundo muy especial. Se interesa por el mundo, el medio ambiente y por la política, pero odia la sopa.
9. Desde 1991 se disputa un Campeonato de Fútbol Femenino y en 2003 la selección de fútbol argentino se clasificó por primera vez para la Copa Mundial.
10. El fútbol fue introducido en Argentina a finales del siglo XIX por inmigrantes británicos.
11. Hoy existen dos ligas nacionales: la Primera División y la Nacional B con dos campeonatos por año.
12. Dice una leyenda que los fundadores del club decidieron tomar la bandera del primer barco que pasara por el río Riachuelo como colores de su club. Fue un barco sueco por eso sus colores son azul y amarillo.
13. El Estadio Monumental pertenece a River Plate y allí juega también la selección nacional. Además, es un lugar para conciertos internacionales de rock y pop.
14. La yerba mate son hojas secas, ligeramente tostadas y picadas o molidas. Para la preparación del mate normalmente se llena la mitad de la calabaza con yerba.
15. El mate frío es una bebida típica para los meses de calor.
16. El barrio de Recoleta es famoso por su cementerio y su centro cultural.
17. En Argentina y Chile la época de la dictadura se caracterizó por la violación sistemática de los derechos humanos: secuestros, arrestos, torturas, extorsiones.
18. Yo, yo he comprendido hasta qué punto callar es importante. Ahora, por las noches, después de comer, mi abuela me lleva siempre a casa de Carlitos, su hermano.
19. Se para, nuevamente. Y nosotros lo abandonamos apenas arrimado a la vereda, y nos perdemos por las calles transversales, sin mirar una sola vez atrás.

8. Anhang

20. Uno de los mayores problemas del país es la desocupación. Hoy el 10,7 por ciento de la población activa está sin trabajo.